Bilal Barakat

"Sorry I forgot your birthday!": Adjusting apparent school participation for survey timing when age is measured in whole years

Article (Published)
(Refereed)

Original Citation:
Barakat, Bilal (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. pp. 300-313. ISSN 07380593

International Journal of Educational Development 49 (2016) 300–313

Contents lists available at ScienceDirect

International Journal of Educational Development

journal homepage: www.elsevier.com/locate/ijedudev

"Sorry I forgot your birthday!": Adjusting apparent school participation for survey timing when age is measured in whole years

Bilal Barakat

Wittgenstein Centre for Demography and Global Human Capital (IIASA, ÖAW/VID, WU), Vienna Institute of Demography (VID), Wohllebengasse 12-14, A-1040 Vienna, Austria

ARTICLE INFO

Article history:
Received 3 April 2013
Received in revised form 28 March 2016
Accepted 30 March 2016
Available online 13 May 2016

Keywords:
Net attendance
Indicators
Household surveys
Measurement error

ABSTRACT

When only whole years of age are recorded in survey data, children who experienced a birthday since the beginning of the school year may appear to be of school-age when they are not, or vice-versa. This creates an error in estimates of school participation indicators based on such data. This issue is well-known in education statistics, and several procedures attempting to correct for this error have been proposed. The present study critiques current practice and demonstrates that its limitations continue to confound educational research and high-stakes policy conclusions: speculative explanations have been proposed for what is actually a measurement artefact. An alternative adjustment strategy is proposed that coherently exploits all available information and explicitly indicates the remaining uncertainty. The application of the method is illustrated by a number of empirical case studies using recent household survey data. These examples demonstrate that the method is feasible, accurate, and that taking survey timing into account can significantly alter how these data are interpreted.

1. Introduction

In a study of education reforms in Mozambique (Fox et al., 2012) it was found in survey data that 75 percent of six-year-olds who were not enrolled in school were reported by their own parents to be 'not of school age'. Given that age six is the official primary school starting age, the study report interprets this curious finding in terms of subjective parental perceptions of school readiness. While these may well form part of the explanation, many of the parents may simply have been stating a fact: their child was six at the time of the survey, but still aged five at the beginning of the school year.

This issue, that some children appear to be of school-age who are not, and vice-versa, arises whenever survey data are collected part-way through a school year. Indicators of school participation, in particular the net attendance ratio (NAR),[1] suffer a distortion as a result. In principle, this effect is straightforward and widely acknowledged. Indeed, when information on birthdays is available (which is the exception rather than the rule), calculating children's ages the beginning of the school year is routine practice in the estimation of education indicators. But handling the issue *when only whole years of age at the time of survey enumeration are available* is markedly less routine.

It is shown here that such whole-year age data (or 'integer age') are still common, and that the error induced by ages unadjusted or insufficiently adjusted for survey timing is potentially large. The NAR is distorted to the tune of ten percentage points in several empirical examples in this study. With respect to the share of children of the official entry age who are attending school, the error reaches 30–40 points in one of the examples presented (Table 5). Since these magnitudes equal or exceed the scale at which policy is assessed, it is unsurprising that such errors confound research and policy conclusions in practice. For example, the true trend in attendance, or differences between groups or countries, may be either obscured or exaggerated, or the relative importance of late entry and drop-out misunderstood. Unfortunately, existing adjustments are at best partial. As a result, they create a false sense of security, so that, as in the example from Mozambique mentioned above, the residual contribution of survey timing to puzzling patterns in apparent school participation is overlooked or misunderstood even after the issue as such has been recognised and purportedly accounted for.

E-mail address: bilal.barakat@oeaw.ac.at.

[1] The NAR is defined as the number of individuals in the official age group attending a given level of education, expressed as a percentage of the total population in that age group. Replacing 'attendance' status with 'enrolment' status in this definition yields the *net enrolment ratio* (NER). The implications of other variants of NER/NAR are discussed in Section 3.6.

http://dx.doi.org/10.1016/j.ijedudev.2016.03.011
0738-0593/

This study is the first that systematically and rigorously seeks to answer the following questions regarding the survey timing effect: What magnitude can this error reach in principle, and how large is it typically in practice? How does it interact with other factors, such as population growth, or drop-out? Why have existing attempts at correcting for this error had only limited success, how can these limitations be overcome to the greatest extent possible, and the remaining uncertainty made explicit?

That these questions remain unresolved is demonstrated by the fact that, while recognizing the existing attempts to adjust for survey timing, UIS (2010, 44) nevertheless recommended further reviews of 'how age reporting affects enrolment [and] survey data' and additional efforts to 'investigate potential ways of [...] adjusting the data'. The present study responds to this call. Through an in-depth review of the literature and an analysis of key education statistics for three country case studies, it explores the relationship between survey timing and integer age data in order to build a better understanding of the age shift phenomenon alluded to above, and other secondary effects. Further, the paper proposes a set of practical adjustments for the NAR that extend the current state-of-the-art in several directions. In particular, these coherently use all the available information, and indicate the remaining uncertainty by providing both lower and strict upper bounds in addition to a 'best guess' estimate.

The practical contribution of this study is to provide insights and methods that allow researchers and policymakers to recognize patterns in apparent educational participation – including differences between groups or countries, trends, and trend changes – that are actually measurement artefacts, and therefore require neither an explanation in terms of educational behaviour nor a policy response.

The following Section 2 substantiates the above claims about the continuing lack of concern for the implications of survey timing in some parts of the literature, and the unsatisfactory nature of existing approaches to accounting for it when only integer ages are known. The gaps in understanding of the potential magnitude of the effect and its interaction with other phenomena (such as drop-out) are addressed in Section 3, before a novel set of adjustment procedures is derived in Section 4. Their empirical application and validation is illustrated in Section 5. The paper concludes with a discussion of the proposed new strategy and some suggestions for future research.

2. A critical review of the state-of-the-art

In the following, 'observed integer age' or 'nominal age' is short for 'age in whole years/age at last birthday measured at the time of survey'. Also, 'the survey timing effect' means the effect of survey timing relative to the school year on educational participation indicators based on nominal age.

The mechanism through which this effect occurs is simple enough, but intuition may be aided further by the graphical representation in Fig. 1. This graph resembles the 'Lexis diagram' in demography, with age on the vertical axis, and time – here: school years – on the horizontal axis. Individuals and cohorts follow diagonal trajectories from bottom-left to top-right, as does the true range of primary-school-age (shaded area), namely those individuals who were aged 6 at the beginning of the school year in this example. By contrast, those *nominally* of school-age at the time of observation are located in the rectangle. The true and nominal school-age populations therefore differ in two 'triangular' areas: erroneously excluded are those at top end of the nominal age range at school start who experienced a birthday (crossed a horizontal line) since; erroneously included are those who entered the nominal entry age since school start. The former are in fact of school-age, while the latter are too young. The key point is that in

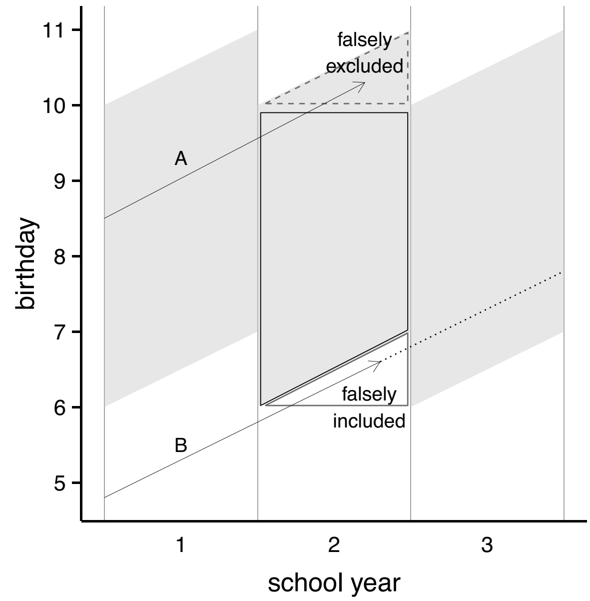

Fig. 1. An educational Lexis diagram.

general, the erroneously excluded are therefore more likely to be attending school than the erroneously included.

2.1. Age recording in household surveys

Given *exact* ages, the correct procedure for correcting the survey timing effect is already well-understood. To see how often this applies, it is worthwhile to briefly review the characteristics of the international household surveys commonly used in global education statistics and development research. Providing a comprehensive review of the wide variety and large number of *national* surveys available (Education Policy and Data Center (EPDC), 2009) is beyond the scope of this paper, but some illustrative examples of specific national surveys are highlighted. Indicators derived from official school register data face a separate set of issues (such as non-alignment between the school year and the reference date for population projections), but typically do not suffer from the survey timing effect that is the present focus. For this reason, the present study focuses on attendance and the NAR (typical for surveys) rather than enrolment on the NER (typical for administrative data).

The main internationally comparable household surveys in developing countries are the Demographic and Health Surveys (DHS) and the Multiple Indicators Cluster Surveys (MICS). Both are nationally-representative, available for many low and middle-income countries for multiple waves three to five years apart, and contain increasingly sophisticated education items. While not without problems,[2] they benefit from large samples and a relatively high degree of comparability across individual surveys. Accordingly, they form a standard data source for international agencies and researchers/policy analysts, complementing or substituting for administrative data. For both DHS and MICS, the

[2] A well-known concern, especially with early DHS waves, is age misreporting. The phenomenon of 'age heaping' at digits 5 and 0, is a general concern with surveys in low-education environments, but the DHS questionnaire structure potentially induces an additional bias. Specifically, the detailed module for children aged five or below may create an incentive for enumerators to record children as being older and thus ineligible for the extra questions (Pullum, 2006). This would typically create a downward bias in attendance indicators both adjusted and unadjusted for survey timing; a potential interaction with their difference requires a separate study.

timing of enumeration relative to the school year varies considerably between surveys even within a single country.

For children between 6 and 14, who are too old for the more detailed 'Child Questionnaire', and too young for the more detailed 'Individual Respondent Questionnaire', the standard DHS household roster collects only integer ages. The same is true for the MICS model questionnaire prior to the introduction of version 4 in the year 2009. The latter also queries birth months, as does the Education Data Survey (EDS) that supplements the DHS in a handful of countries.[3]

Note that the age shift underlying the survey timing effect is not avoided by querying 'attendance at any time' during a given school year instead of strictly 'current attendance'. Indeed, the former underlies most of the empirical examples in this report, including some with a significant effect. While 'any time' measures are theoretically less sensitive to the interaction of survey timing with contemporaneous *drop-out* (cf. Section 3), the practical difference with 'current attendance' is small unless drop-out occurs mostly during the school year, rather than between them. Indeed, 'any time' attendance referencing a specific school year potentially even increases the risk of extreme survey timing distortion, if such data are collected more than a year after the beginning of said school year and left unadjusted. For example, if in December 2011 we query whether children ever attended school during a school year running from September 2010 to June 2011, then – assuming an official entry age of 6 – *none* of the observed 6-year-olds were actually of school-age at the time, as well as many of the 7-year-olds even. Such long delays do in fact occur in DHS data, for instance. By itself, information on 'age at school entry' does not resolve the issue either; it needs to be combined with the year of entry and/or data on repetitions in order to yield the age at the start of the school year for students other than new entrants into the first grade.

Outside of DHS and MICS, too, integer age data remains far from exceptional. With reference to the present case study countries, the Kenya Integrated Household Budget Survey (KIHBS) of 2006 contains only integer age, as does the 2012 SUSENAS in Indonesia. More broadly, among the most recent datasets available for Sub-Saharan African countries among the International Public Use Microdata Series (IPUMS), fewer than half include birth months, and many of those contain more missing values than data entries. Unsurprisingly, the situation is worse for historical data. And while the ongoing longitudinal YoungLives survey (Boyden, 2014), with its highly-detailed education module, naturally records exact birth dates for its panel members specifically, only integer ages are recorded for other children in the household; since the panel is restricted to certain cohorts, the calculation of NARs must be based on the latter. The World Bank Microdata Library also contains recent surveys with integer age data only, both under the rubric of 'Living Standards Measurement' and 'Impact Evaluation'. As a non-exclusive example of the latter, consider the Education Outcomes National Panel Survey (NPS) 2002–2008[4]: despite being specifically designed as an impact evaluation, in other words, requiring maximal comparability between the baseline and follow-up study, the two waves were conducted at very different times in the school year and recorded only integer ages.

In sum, there is a large legacy of educational data from household surveys that measured nominal ages only, and – high-profile counterexamples notwithstanding – this remains a common practise even for newly-designed questionnaires.

2.2. The survey timing effect in the literature

Turning from integer age data as such to their appraisal, several stances can be identified in the literature.

In some cases, the potential distortion resulting from survey timing goes without mention. This omission occurs even in the specific context of discussing measurement issues relating to school attendance, including both original substantive research (Carr-Hill, 2012)[5] and methodological overviews in seemingly comprehensive and authoritative handbooks (e.g., Glewwe, 2000; Glewwe and Kremer, 2006; Orazem and King, 2007). Strikingly, the chapter 'Schooling in Developing Countries' in the voluminous *Handbook of Development Economics* (Orazem and King, 2007) avoids touching upon the issue in a 13-page section dedicated to 'Measurement Matters' (p. 3536), despite launching the discussion on measuring enrolment status with a *different* point 'depending on when the questionnaire is applied', and another relating to the 'reference period' for school attendance.

Since the existence of the survey timing effect as such is well-understood, such omissions presumably indicate an expectation that its magnitude and practical implications are negligible. This assumption may be shared even when the survey timing effect is indeed acknowledged. The most current DHS guidelines note the effect and announce the planned adoption of the adjustment procedure followed in MICS (ICF International, 2013) (see below). Nevertheless the past indicator values published in the online *DHS StatCompiler* database remain unadjusted at the time of writing.[6] Indeed, all of the final reports for individual DHS surveys analysed for the present study simply remain silent on the issue. Elsewhere, the assumption that survey timing has at most a moderate impact is made explicit. The magnitude of the effect is noted to generally result in 'a few percentage point' difference (Education Policy and Data Center (EPDC), 2009, 16), for example. A UIS (2010) report on sources of disagreement between administrative and household survey data states that survey timing cannot account for the differences under discussion, which are in the range of 5–20 points in estimates of overage enrolment. A UNICEF (2014, 11) report implies the inaccuracy will not be considerable, unless the survey was carried out in the latter half of the school year. None of these documents present a quantitative argument for their claims. This is unfortunate, because both theoretical considerations and empirical estimates presented in this study show these claims to be overly optimistic.

In methodological guides, mention of the survey timing effect is not necessarily accompanied by concrete operational advice on how to correct for it. The World Bank guide for its *ADePT Edu* data platform, for example, limits itself to the generic exhortation that the effect 'should be taken into consideration when interpreting education indicators' (Porta et al., 2011, 28), which offers no advancement over the similar formulation employed seven years earlier by UIS (2004, 32). Elsewhere, practical adjustments have indeed been proposed. The necessity of examining these procedures and their limitations in detail, as the following Section 2.3 does, is highlighted by their rather uncritical use in the literature.

For example, EDOREN (2014) acknowledge as a 'vital' problem [p. ii] that surveys 'are collected at different times in the school

[3] However, foreshadowing the cavalier treatment of survey timing documented further below, the report on the 2010 Nigeria EDS, for example, elects to not actually take advantage of the detailed age information in the calculation of attendance ratios, instead relying on unadjusted nominal integer age data in the matched 2008 DHS, without discussion of the implications (National Population Commission (Nigeria) and RTI International, 2011, 54).

[4] Reference ID: MOZ_2002_NPS_v01_M.

[5] Not every study has to discuss every issue, of course. What arguably makes it an omission not to mention survey timing in this context is that the effect contributes not a random error, but a systematic bias in the estimated share of out-of-school children in the opposite direction to that being discussed.

[6] While undocumented, this can be established by reproducing these values without performing any age adjustment.

year'; they further note the difference between the 2007 and 2011 MICS in Nigeria regarding the methodological treatment of the issue, and that the DHS underwent no adjustment. Nevertheless, the report discusses inconsistencies between indicators based on DHS, MICS, and administrative data at face value, without considering survey timing among the 'possible reasons for discrepancies' [p. 38]. Moreover, the difference between the NAR in MICS reports and the values re-calculated from micro-data are attributed to diverging indicator definitions, when for MICS 2007 it is actually largely accounted for by the age adjustment included in one but not the other.

In the case of the Mozambique NPS mentioned above, having created the conditions for survey timing to induce an error, the issue is raised as an afterthought in an appendix to a technical document (World Bank and Instituto Nacional de Estatistica Moçambique, 2008). There, an ad-hoc adjustment of a whole year is applied to observed ages as the supposedly 'best approximation', and the limited comparability with the 2008 MICS is noted, without recognising that the latter actually analyses similarly adjusted ages. The age shift then receives no further discussion in the actual study report (Fox et al., 2012), despite the fact that in several figures (specifically 3.6 and 3.7) the difference between 2003 and 2008 data can be straightforwardly interpreted as a sideways shift of the curves along the age axis (and despite the fact, already noted in the introduction, that the explanation in terms of the survey timing effect fits better with the justification given by the students' own parents).

Such examples illustrate a tendency to assume that simply because *some* adjustment has been applied, survey timing can no longer explain remaining discrepancies. At the very least, this ignores that for the most widespread approaches based on a mid-school-year threshold, the fact that an adjustment *process* was followed does not necessarily mean that any adjustment was actually performed.

2.3. Existing adjustments

A number of quite distinct approaches to adjusting for the survey timing effect have been proposed over the time, which are discussed briefly below.

2.3.1. Proportional adjustment

An early attempt is documented by UNESCO Division of Statistics (1997), for the general situation of wanting to include only parts of an observed age-group in the calculation of an indicator. In the given example, where some children attend primary school from age 6 to 9, and others from 7 to 10, they suggest including the enrolments at ages 6 and 10 in the numerator, but only half of each age group in the denominator. Unfortunately, this is incoherent at the individual level: in the example, 10-year-olds are effectively assumed to contribute more enrolments than their number among the pool of children *eligible* for enrolment. In any case, this approach does not appear to have been adopted more widely since its initial introduction.

A procedure for adjusting under-age and over-age enrolment for survey timing presented by EPDC (2009) attempts to formalise the notion that the proportion of students with a birthday since the school year beginning equals the fraction of the school year elapsed. Said percentage of seemingly over-age students are then assumed to actually be on-time, and of seemingly on-time students to actually be under-age. A proportional adjustment implied by randomly distributed birthdays is similar in spirit to one adjustment proposed in this study. However, the above procedure is not a precedent, because it is misspecified: for it to recover the correct values is not merely unlikely, but strictly impossible. The intuition is that the procedure conflates randomness of birthdays 'looking forward' from the initial timeliness status on the one hand with randomness 'looking back' from the status observed during the school year on the other. A formal proof outline is included in the appendix. For a straightforward example of failure in a special case, note that if under-age enrolment is truly absent, the stated procedure cannot reconstruct this fact.

The correct adjustment for age-appropriate enrolment is not pursued any further. The age shift in observed timeliness, and its interaction with drop-out, is considerably more complex than with enrolment/attendance as such. Moreover, early and late entry likely vary by birth month, which determines who is perceived as 'almost' or 'barely' of school age. These complications merit a separate treatment. Anyhow, apart from being misspecified, the EPDC (2009) adjustment for age-appropriate enrolment has not seen adoption as a method for adjusting attendance itself. For the latter, recent publications by EPDC implement alternative strategies.

2.3.2. Alternative data with exact birth date

In Omoeva et al. (2013), the issue of survey timing is sidestepped by exploiting the exact birth dates available in DHS for sampled women's *own* children. Locating these among household members allows for matching their attendance status to their true age. Unfortunately, this approach creates a new bias. Orphans, but also foster children, do not appear in the household listing of their mothers, and cannot be matched to birth dates. But orphans' schooling may differ significantly from that of children living with their mothers (Case et al., 2004), although not necessarily so (Smiley et al., 2012). Their share is non-negligible in many developing countries, 20 percent or more not being confined to isolated cases. If the unmatched children are omitted, this approach trades one error for another without assessing the magnitude of either, an unsatisfactory situation. Retaining unmatched children leaves unresolved the question of how to adjust their ages.

2.3.3. Discrete one-year adjustment

Another class of adjustments (labelled 'binary adjustments' here) rounds the elapsed fraction of the school year to the nearest integer, and subtracts this number from nominal ages to more closely approximate the exact age at school start. Notably, this approach is not necessarily considered a complete solution to the problem even by its proponents, who recognise that nevertheless, 'additional studies need to be conducted to better understand the issue' (UIS, 2010, p. 42) of age reporting and its implications.

One variant applies the same binary adjustment to all children depending on whether the *median* observation was made in the first or second half of the school year (UNICEF and UNESCO Institute for Statistics, 2015). This procedure has been the method of choice at the UNESCO Institute for Statistics (UIS) for some time (UIS, 2005); it is also current practice for reports by the Global Initiative on Out-of-School Children (UNICEF, 2012, 2014), and was followed in MICS3. An alternative is to apply a binary adjustment only to those specific ages that were recorded in the second half of the school year. This variant underlies the MICS4 surveys (for children where the exact age is unknown), and has now been adopted by DHS (ICF International, 2013, 24) (but was evidently not been applied to the DHS *StatCompiler* database retrospectively). The details of this procedure have enjoyed limited dissemination, being documented only in the MICS data processing syntax files, not the reports, leading even a highly systematic data review (EDOREN, 2014, 13) to conclude that MICS adjusts ages 'without providing further details on the methodology'.

Median and individual binary adjustment share some inherent weaknesses. First, the survey timing effect can be non-negligible even if the survey was conducted in the first half of the school year (Table 4). Second, determining a school-year's exact mid-point can be difficult (Section 5.3). Third, with respect to *comparisons* over space and time, real changes are smoothed out or spurious discrepancies introduced if some cases are rounded up and others down, creating an undesirable trade-off between consistent *procedures* and consistent *results* (again, Section 5.3). Fourth, both binary adjustments discard known information about the approximate distribution of true ages (Section 3).

The median binary adjustment faces a number of additional challenges, both conceptual and practical. Firstly, the informal notion of the (chronologically) 'median observation' is underdetermined: The median *household*, or the median school-age *individual*? Taking sample weights into account, or not? Replicability requires either a general consensus or meticulous documentation of these choices. Secondly, and more profoundly, it is conceptually incoherent to make the adjustment to a given respondent's age depend on when *other* respondents were interviewed. This latter contingent fact simply carries no information regarding the former's true age.[7] Thirdly, the ambiguity of exactly *which* median to use emerges as a concrete practical problem when multiple levels of aggregation are considered. In particular, the median binary adjustment is inconsistent with respect to disaggregating indicators by subnational units (Section 5.3).

There is a false sense of security in the accuracy of data where the age shift has been adjusted but only *partially* (as both kinds of binary adjustment do). Recall that in EDOREN (2014) above, it is explicitly recognised that unlike MICS 2011, the 2010 DHS in Nigeria does not account for survey timing; yet it is claimed that '[i]t is not clear what causes the discrepancies' [p. 6], even though adjusting either both or neither largely removes the disagreement. As another example, the Nigeria MICS 2007 report notes as a table footnote in the appendix that education indicators are based on 'estimated age as of the beginning of the school year', but in the discussion of the results makes no mention of whether patterns remarked upon may be related to limitations of this estimation. This despite the fact that the survey was approximately half-way through the school year, the worst case for the binary adjustment that was applied.

In summary, none of the existing approaches, spread across technical reports, is universally established among educational statisticians, much less among the wider research community. The problem is not simply limited dissemination. All procedures identified above are relatively simplistic and suffer from either limited applicability (i.e. they cannot be used in all data settings) or limited effectiveness (i.e. they do not actually correct the error reliably).

What the existing literature and current practice do not offer, when participation ratios must be based on integer age data, are: (a) a quantitative assessment of the potential magnitude of the error induced by survey timing, and a systematic analysis of its interaction with drop-out, population growth, and question design, and (b), an operational adjustment that is coherent, uses all available information, but acknowledges the remaining uncertainty. This study contributes to closing these gaps.

[7] Arguably, the situation is even worse. Consider basing the adjustment on whether the majority of *attempted* interviews occurred before or after the mid-year threshold. Presumably, this would be considered to be poorly justified. But the conceptual difference to current practice regarding the median binary adjustment is not entirely clear-cut: realised interviews may be discarded during quality assurance for failing consistency checks. At worst, including or omitting such a case might determine the timing of the median interview relative to the threshold. In theory, therefore, the adjustment applied to some respondent's age following the median binary procedure potentially depends not only on *when* other individuals were questioned, but on *what* they responded to *unrelated* questions.

3. Primary and secondary effects of survey timing and integer age measurement

This section provides a comprehensive overview of the different ways in which survey timing affects apparent educational participation through *secondary* and indirect effects. It continues to refer to Fig. 1.

3.1. Magnitude of the primary age-shift effect

The primary consequence has already been described: the later in the school year, the greater the share of the erroneously included and excluded. The point is that this exchange is not neutral with respect to apparent participation. If all entry is on time and progression universal, children *not yet attending* have erroneously replaced children that *are attending* in the assumed pool of school-age children. As a result, the apparent attendance ratio is lowered.

Two implications for the error magnitude are clear from Fig. 1. First, if births occur uniformly over months, the error grows linearly as the school year progresses. Second, in the worst case, just before the school year ends, an entire single-year age group (or grade) has been swapped: all nominally 6-year-olds were only 5 at school start, and all starting school at the top of the school-age range have since aged out of the range. If none of the former are enrolled/attending, but all of the latter, the maximal error in the overall participation ratio can therefore reach $1/n$, where n is the number of grades. Concretely, for common variants of the primary cycle with 4 to 8 grades, the nominally observed attendance ratio could in principle be lowered by 12.5–25 points. As noted in the literature review, the potential for such an error magnitude tends to be underestimated. Just before the school year mid-point, when binary adjustments are not yet 'triggered', almost half of the error already pertains. Such 6–12 points in the worst case clearly already hold considerable potential for distorting policy conclusions. For example, a 'true' increase of 5 points in attendance could be mistaken for a 5 point decline.

Moreover, it is intuitively clear that an indicator such as the *Net Intake Rate* (NIR), defined as the share of entry-age children attending the first grade, is even more strongly affected. Indeed, for the NIR there is no theoretical bound whatever on the error: in the absence of early entry, a survey measuring nominal age at the very end of the school year would result in an estimated NIR of zero, even if timely entry were in fact universal. However, as with measures of age-appropriate enrolment, the additional complexity of considering participation by grade requires a separate treatment.

A number of secondary effects have received little to no attention in the literature, but are mostly intuitive with reference to the diagram.

3.2. Early entry

all else being equal, *the more early entry, the smaller the survey timing effect*. This potentially fully offsets the main effect. If all of those prematurely included are already attending anyhow, just as the erroneously excluded, then replacing attending children with other attending children leaves the attendance ratio constant.

3.3. Cumulative drop-out

Conversely, all else being equal, *the higher cumulative drop-out, the smaller the survey timing effect*. As above, if none of the erroneously excluded are attending, like the erroneously included, then replacing non-attending children with other non-attending children has no effect.

A combination of high early entry and high cumulative drop-out implies the nominal NAR may theoretically *overestimate* the NAR late in the school year. In practice, a combination of *late* entry and high drop-out is more typical (Wils, 2004).

3.4. Population growth

Population growth, or more precisely: cohort-on-cohort growth,[8] affects the true participation ratio directly (Barakat et al., 2013). But in addition, all else being equal, *the higher cohort-on-cohort population growth, the greater the survey timing effect*. The erroneously excluded, more likely attending, are replaced by a larger number of erroneously included, assumed less likely attending. Consequently, the attendance of those truly of primary school age is diluted more. Given a constant growth rate, the size difference will increase for birth cohorts further apart. A corollary is that *the longer the primary school cycle, the larger the cohort-on-cohort population growth amplification of the survey timing effect*. However, simulations show that for typical population growth rates, the population growth effect is an order-of-magnitude smaller than the primary effect. Nevertheless, given extreme population volatility resulting from crisis-induced migration, for example, the included and excluded age groups could conceivably differ sufficiently in size to increase the primary effect by several percentage points.

3.5. Exact reference date

So far, equating the 'true' age with the age at school start has assumed that the reference date for determining eligibility for schooling coincides with the beginning of the academic year. Deviations from this are not uncommon, and there are examples both of reference dates before and after (often December 31) the beginning of the school year. For example, in Germany cut-off dates vary between states by half a year, from June 30 to December 31, and the equivalent exact entry ages at school start from 6 years 2 months to 5 years 8 months (bildungsserver.de, 2014) within a single country. Ideally, age adjustments should therefore be based on survey timing relative to the reference date, rather than relative to the beginning of the school year. If the latter is used for simplicity, the resulting error is equivalent to the error if a corresponding survey timing shift relative to school start remained uncorrected. An important subtlety is that in the case of a true reference date after school start, the direction of all the secondary effects mentioned above is *reversed*; a consequence of the fact that in this case, in contrast to the general pattern, some older children are erroneously included and some younger children erroneously excluded.

3.6. Indicator definition

Over time, the standard definition of 'the primary NER' has shifted. This is also relevant for understanding the NAR, which differs only in terms of the way educational participation is defined (actual attendance instead of mere administrative enrolment), but is otherwise structurally equivalent. In either case, the denominator is always the number of primary-school-age children. Originally, only primary-school-age children strictly in *primary* grades entered the numerator. However, under this strict definition, the NER (as well as the NAR) is diminished in the presence of children who enter early and progress more rapidly. These children may already have *completed* primary school while still in the eligible age bracket, and either entered the secondary level or left school. Either way, they are not counted as being 'in primary school' under the original strict definition of the NER/NAR. Including such children yields the 'adjusted' NER/NAR.[9] Since older children erroneously excluded based on nominal ages are more likely to count as 'attending' for the adjusted NAR than for the strict NAR, this implies that *the 'adjusted NAR' is more sensitive to the survey timing effect than the strict NAR*. Because the term 'adjusted NAR' may cause confusion with the unrelated 'adjustment' of ages, the remainder of the manuscript omits the qualifier, so that 'NAR' is understood to mean the 'adjusted NAR', unless noted otherwise.

The following points only apply to strictly 'current attendance', rather than attendance at any time during a given school year.

3.7. Contemporaneous drop-out

Independently of the age shift, *the higher contemporaneous drop-out, the larger the survey timing effect on current attendance*, simply because later in the school year, a larger share of eventual drop-outs has already left. However, drop-out over the course of the school year is, of course, connected to cumulative drop-out. Since the latter acts in the opposite direction (see above), a natural question is: which dominates? Actually, if drop-out were uniform over the school year and across grades and stable over time, they would cancel each other out exactly. From a demographic perspective, this follows directly from the equivalence of cohort and period perspectives under constant rates.[10] In reality, of course, drop-out may not occur evenly distributed over the school year, or across grades.

3.8. Treatment of the break between school years

Between school years, the 'current attendance' status of prospective new entrants is unclear. Instructions may differ between surveys (compare, for example, the contradictory instructions for the Indonesia 2010 and Jamaica 2001 censuses on this point IPUMS, 2014), or leave the status of prospective entrants unspecified. A related issue can be subsumed under this rubric: some school courses, or at least their final grade, may not last a whole year (UIS, 2012) (typically at the secondary, not primary, level). In either case, the problem is that graduates have departed before new entrants are in physical attendance. Not counting prospective entrants as attending means the apparent

[8] The key question is, with reference to the example in Fig. 1, whether the total number of 10-year-olds (whether attending or not) is considerably greater or smaller than the total number of 6-year olds. While it is likely to be smaller if overall population growth is high, this is not necessarily the case: even after fertility has declined and each birth cohort is smaller than the preceding one, the total population could still be growing, if adults are surviving for much longer, for example.

[9] Unfortunately, departing from the strict definition creates some operational ambiguity that can impede the exact replication of published figures. For example, the data processing code underlying MICS tabulations includes children who left school after completing primary school in the numerator for the adjusted NAR, but not the (few) children who completed primary, entered secondary, and *then* left school. These should in principle be included, as should those who graduated from primary school, but are now enrolled in non-secondary post-primary education. Restricting the primary graduation criterion to children of normal age for the last primary grade is also questionable.

[10] For a derivation from first principles, note on the one hand that the students contributing to the primary effect are those in the final year. Let annual cumulated drop-out be $100r$%, and let $s = 1 - r$, and there be n grades. Then, only s^n of the oldest entry cohort are left at the end of the final year, so counting them (exact age) or not (whole years age) makes a difference, in absolute share, of: $(1 - s^n/n)$. On the other hand, at the beginning of the school year, the total student population is $(1/n)\sum_{i=0}^{n-1} s^i$, and a fraction $r = 1 - s$ of these will drop out by the end of the year. So in absolute terms, the total drop-out in all grades together will likewise be

$$\frac{1}{n}(1-s)\sum_{i=0}^{n-1} s^i = \frac{1}{n}\left(\sum_{0}^{n-1} s^i - \sum_{1}^{n} s^i\right) = \frac{1-s^n}{n}.$$

participation remains at a constant reduced level throughout the break; counting them means participation increases monotonically to peak at the beginning of the new school year, as 'attending' prospective entrants age into, and 'non-attending' graduates age out of, the school-age bracket.

4. A set of adjustments

The adjustment procedures proposed here correct the primary age shift effect, but not contemporaneous drop-out and other secondary effects. They all merely re-assign age, but leave the attendance variable itself untouched, ensuring that total attendance across all ages always remains the same. These procedures arise as natural extensions of current practice.

4.1. Random/proportional age assignment

A straightforward refinement of the median binary adjustment is to calculate a weighted average of the NAR with and without shifting all ages, weighted by the elapsed fraction of the school year. Alternatively, starting from the individual binary adjustment, one could consider data collected over a short time period straddling the school-year mid-point, adjusting half the ages (but not the other half), resulting in a more accurate overall estimate than if all were on the same side of the threshold. Again, this suggests replacing the single threshold with an approach where a proportion of ages is adjusted matching the elapsed fraction of the school year.

Consider Table 1a, framing the survey timing effect as a problem of imputing missing information. For simplicity, the presentation is for individuals of nominal entry age. The younger individuals had a birthday since school started and are not actually of school-age, while the older individuals are. The margins of the table are known (up to non-uniform birth distribution and sampling error), but both binary adjustment procedures currently in use contradict this knowledge by assuming one of the columns to be all zeroes. The arguments above suggest assigning the product of the marginal shares to each cell instead.

4.2. Age assignment ordered by attendance status

This random assignment is likely to *underestimate* the true NAR: if children of entry age at school start are more likely to enter than those too young, then the odds for belonging to the older group are higher for a child attending than for a child not attending at the time of survey.

Exaggerating this link in order to arrive at a definitive *overestimate* implies the assignment in Table 1c. This formalises the intuitive process of assigning individuals to the older age starting with those attending. For example, if mid-way through the school year, half the children of nominal entry age are attending school, then all attending are assumed to be truly of school-age, with no early entry. The more complicated expressions in the table allow for the share attending to exceed or fall short of the elapsed fraction of the school year. In the following, this approach will be called the strictly 'ordered' adjustment.

In application, a preparatory step for observations collected more than a year after the beginning of the reference school year is to reduce both age and p by 1, forcing $p \in [0, 1]$. Both the random and the ordered assignment are carried out for all ages within the school-age bracket, including one year above. Instead of an actual randomised assignment according to assigned cell probabilities, the sample weights are re-distributed instead, avoiding the need to average across many replications. For the ordered assignment, the 'direction' is reversed for ages in the upper half of the school-age bracket, meaning those still attending are preferentially assigned to be younger. The rationale is that older individuals are more likely to have already graduated or dropped out. After performing the assignments for all observed integer ages and observation periods, the NAR can be estimated on the adjusted data. For a point estimate, the mid-point of the interval between the estimated bounds would be a convenient convention, reflecting the fact that attendance is likely associated with true age, but not perfectly so.

4.3. Regression on interview dates

Alternatively, we can empirically *estimate* the strength of this association if the data collection is spread over time. Suppose a share r_6^{true} of then 6-year-olds entered school at the start, and r_5^{true} of then 5-year-olds. Then a fraction p into the school year, the observed attendance among nominally 6-year-olds is the weighted average

$$r_6^{obs}(p) = p \cdot r_5^{true} + (1-p) \cdot r_6^{true}.$$

Given $r_6^{obs}(p)$ for varying p, we can statistically estimate r_5^{true} and r_6^{true}.

A simple linear regression is inadequate, because the intercepts of the regression line with $p = 0$ and $p = 1$ represent age-specific attendance ratios (ASAR) bound by the interval [0, 1]. Since the relationship is approximately linear,[11] logistic regression is also inappropriate. Instead, the ASARs that parameterise the mixture model can be estimated directly by Maximum Likelihood.

While this yields a 'predicted values' for the ASARs at school start, these are not aggregated directly to the NAR. Like any statistical model, the fit may be imperfect; using the 'predicted' ASARs therefore risks creating attendance 'out of thin air'. To avoid this, predicted ASARs are instead used indirectly to re-weight the observed data. Formally, the posterior odds for an individual with observed integer age a of having been aged $a-1$ at school start, conditional on their attendance status, follow

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds}.$$

In particular, this means:

$$\frac{P(\text{age}^{SYS} = 5 | \text{attending})}{P(\text{age}^{SYS} = 6 | \text{attending})} = \frac{P(\textit{attending} | \text{age}^{SYS} = 5)}{P(\textit{attending} | \text{age}^{SYS} = 6)} \times \frac{P(\text{age}^{SYS} = 5)}{P(\text{age}^{SYS} = 6)},$$

and conversely for the non-attending, where age^{SYS} is the age at school year start. The prior odds are approximately $p/1-p$ with p the elapsed fraction of the school year, and the likelihood ratio derives from the regression. So at each observation time, the odds of a birthday since school start are updated for the attending and non-attending separately, and the sample weights for each group distributed between the two possible true ages according to the updated odds.

The regression-based NAR estimate is not recommended as a universal standard approach; like any statistical estimation, it can fail to capture the underlying relationship for any number of reasons. Notably, it suffers bias if data collection timing correlates with predictors of attendance. For example, fieldwork may be spread over an extended period in order to visit different geographical areas sequentially. Also, even when timing and attendance covariates are unrelated and the adjusted NAR itself unbiased, studying the effect of the covariates on attendance should not be performed post-hoc on the age-adjusted data; instead, the covariates should already enter the age-adjustment

[11] Linearity holds if attendance depends only on integer age at school start without variation by birth month. Simulations show that plausible magnitudes of the latter induce non-linearity that is marginal relative to the estimated slope.

Table 1
The fundamental problem relating to integer age measurement when some fraction e of the school year has passed: the approximate share of children who have had a birthday since the beginning of the school year is known to be approximately e, and the ratio attending r is known, but not how the two relate. The random (or proportional) assignment: attending and non-attending children are assumed to be equally likely to have had a birthday since the beginning of the school year. The "ordered" assignment, which assumes that those attending are strictly more likely to have already been of school-age at the beginning of the school year. N.b. either the first or the second option applies in all the minimum/maximum selections at the same time, depending on whether $r<(1-e)$.

(a) The assignment problem.

	young	old	**total**
attending	?	?	**r**
not attending	?	?	**1-r**
total	**p**	**1-p**	**1**

(b) Random assignment

	young	old	**total**
attending	$p \cdot r$	$(1-p) \cdot r$	**r**
not attending	$p \cdot (1-r)$	$(1-p) \cdot (1-r)$	**1-r**
total	**p**	**1-p**	**1**

(c) Ordered assignment

	young	old	**total**
attending	$max(0, r-(1-p))$	$min(r, 1-p)$	**r**
not attending	$min(p, 1-r)$	$max((1-p)-r, 0)$	**1-r**
total	**p**	**1-p**	**1**

regression itself. Moreover, weights reflect sampling strata potentially based on characteristics such as region, urban/rural residence, or even ethnicity. This creates subtle complications. Specifically, the sampling weights may correlate with the age distribution, conditional on attendance. This issue requires further research. Certainly it is prudent to check whether the conditions for a successful regression-based adjustment are met on a case-by-case basis.

4.4. A model-free descriptive alternative: truncated NAR

The opposite applies to a simple, assumption-free indicator that is robust to the survey timing effect instead of attempting to correct it, namely calculating the NAR on a truncated age range that omits the lowest and highest age of the nominal range. If the observed nominal age is in this range, the true age is certainly in the full school-age bracket. Barring late entry and drop-out, this truncated NAR will be unaffected by survey timing. Under less ideal circumstances, it becomes less sensitive to late entry and more sensitive to cumulative drop-out over the course of the school year. For strictly current attendance, it also tracks contemporaneous drop-out. Nevertheless, for trends over time, comparing the truncated NAR to the ASAR at the nominal entry age may indicate whether survey timing is distorting the regular NAR. The potential benefit rests in doing this in purely descriptive terms on the nominal age data, without having to justify the assumptions behind an adjustment. An example application is included in Section 5.4.

5. Empirical case studies

In the following country case studies, these ideas are applied to empirical data. The first two case studies serve to illustrate and further elaborate on the methods; the third is a demonstration that the adjustment can completely change the interpretation of a policy analysis. First, the functioning of the regression-based adjustment discussed above is illustrated with data from Indonesia. Second, a consideration of national and state-level data from Nigeria shows why the whole-year age shift conditional on the timing of the median interview is unsatisfactory as a general approach. Third, a recent policy analysis of trends in primary attendance in Sub-Saharan Africa, and specifically Kenya and Rwanda, is critiqued by re-evaluating the face value evidence when the age shift is taken into account. Each case study presents estimates specific to the point under discussion; in addition all estimates for the NAR for the country as a whole are included in Table 4, with the following labels (Table 2):

5.1. Data

Two datasets are used in addition to standard MICS and DHS surveys, whose general characteristics have already been outlined in Section 2.

The Indonesian case study of the regression-based age adjustment further utilises two additional household surveys, namely the Indonesia Family Life Survey (IFLS) (Strauss et al., 2009) and the national socio-economic household survey SUSENAS (Statistics Indonesia (BPS), 2012). The IFLS (RAND Corporation, 2014) is a longitudinal household survey representative of 13 Indonesian provinces containing about 83 percent of the national population. The most recent wave 4 was conducted by RAND in cooperation with the local Center for Population and Policy Studies (CPPS) of the University of Gadjah Mada and Survey METRE. It is included in the analysis because it contains exact birth dates for all children, thus allowing for a validation of the regression estimate. The SUSENAS (Statistics Indonesia (BPS), 2012) is included as a third, independent validation of the estimates. It is a large-scale multi-purpose survey that has been collected by the national statistical office of Indonesia with increasing frequency and is currently collected quarterly with a rotating panel.

Basic descriptive information on all surveys, including the timing of the data collection and the primary school calendar, as well as characteristics of the age and attendance queries, are collected in Table 3.

Table 2
Adjustments of integer observed age for survey timing.

Label	Adjustment	Definition	Original
NM	Nominal (= no adjustment)	Self-explanatory	
MB	Median binary adjustment	Section 2.3.3	
IB	Individual binary adjustment	Section 2.3.3	
RD	Random assignment	Section 4.1	*
RG	Regression-based adjustment	Section 4.3	*
OD	Ordered assignment	Section 4.2	*

5.2. Indonesia: demonstrating the regression-based adjustment

Indonesia is an appropriate case study not only because of its size and dynamic educational development. It also offers the availability of multiple large-scale household surveys that can be compared, including the IFLS that records exact birthdays, and the recent 2012 DHS and 2012 SUSENAS. As mentioned, the latter is particularly interesting because it provides the template for several other surveys in Indonesia. Table 5 displays the estimated ASAR for the children aged 6 years at the beginning of the school year.

The regression on IFLS data is shown in Fig. 2a. Note that the intersections with the lines $p = 0$ and $p = 1$ yield the predicted ASAR for ages 6 and 5 respectively at school start. Given the exact birth dates in IFLS, the age at the beginning of the school year is known, but used exclusively to validate the estimates. Despite considerable variability in the outcomes, these true values are predicted well.

The IFLS are almost ideal for the regression-based adjustment; not only can it be validated against exact birthdays, also the enumeration period covers a large proportion of a single school year, and exact interview dates are known. The following two analyses are more challenging. In SUSENAS, the data contain only two approximate enumeration times, but far apart. In the DHS, enumeration was concentrated towards the very end of the reference school year.

Data collection for the 2012 SUSENAS occurred in two short quarterly waves lasting three weeks each. For each observation, only the corresponding wave is known. Despite this limited timing information that forces a regression on only two points, the implied ASARs are highly consistent with the IFLS estimate (Fig. 2b). Technically, the two enumeration periods refer to different school years, but in the presence of stable rates over time the difference is minimal.

The 2012 DHS is shown in Fig. 2c. Unsurprisingly, the regression on nominally-6-year-olds concentrated towards the school year end yields a poor estimate of the intersection with $p = 0$ at the other end. However, the final estimate of the ASAR at age 6 at school start is determined by two groups: nominal 6-year-olds whose age is kept, and nominal 7-year-olds whose age is shifted. In the previous two regression examples, only the former were shown to illustrate the basic principle most clearly and because the two estimates essentially coincide. By contrast, here the ASAR estimated for 7-year-olds at the end of the school year, i.e. the intersection of the upper regression line with $p = 1$, is more accurate.

Indeed, the latter dominates the overall ASAR estimate for 6-year-olds at school start (cf. Table 5). In this sense, the procedure is 'self-weighting': most members of an adjusted age group, say age a, come from the nominal age group which allows for the most accurate estimate of the true ASAR at age a. For a survey early in the school year, the intercept at $p = 0$ is close to the data, and well-estimated. This intercept is the ASAR at true age a for the regression on children of nominal age a, who also make up the bulk of children assigned true age a. Conversely, for a survey late in the school year, the intersection at $p = 1$ is well-estimated; it corresponds to the ASAR at true age $a - 1$ for the regression on children of nominal age a, who also make up the bulk of the true age group $a - 1$.

5.3. Nigeria: a critique of binary adjustments

In Nigeria, the most recent MICS, in 2007 and 2011, were incidentally both conducted approximately half-way through the school year. This represents the worst case for binary adjustments. Just before the mid-point, the maximal unadjusted error remains. Just after the mid-point, the error in the adjusted ages is just as large, merely in the opposite direction (though the NAR estimate's accuracy still benefits in the presence of late entry and delayed progress). Unfortunately, Nigeria's population size leverages the residual error half-way through the year to a significant level.

The 2011 MICS actually records birth months. The estimate using this information (70.1 percent) validates the regression-based adjustment (72.1 percent), which gives a more accurate estimate than any of the currently practised methods (Table 4); in fact, the true fit is even better, because the birth month is missing in fully 14 percent of the relevant observations, and MICS applies the IB procedure to these cases, so the 70.1 is known to be a slight underestimate. The mid-point between the RD and OD estimates, at 70.9, is therefore likely to be very accurate indeed. Conversely, such close estimates from three independent approaches is encouraging news with respect to the quality of the birth month data.

Even taking the value of 70.1 at face value in order to be conservative, the estimation error is 0.8 points for the interval mid-point and 2 points for the RG estimate, compared to 4.4 and 4.1 points for the unadjusted and IB procedures. In terms of the

Table 3
Data sources for country case studies.

Country	Survey	Year	School age	Sample size[a]	Enumeration period	School year	School start
Nigeria	MICS3	2007	6–11	33,403	Mar–Apr 2007	2006/07	Sep
Nigeria	MICS4	2011	6–11	36,272	Feb–Mar 2011	2010/11	Sep
Indonesia	IFLS	2007	6–11[b]	7,375	Nov 2007–Apr 2008	2007/08	July
Indonesia	SUSENAS	2012	6–11[b]	96,827	Mar, Sep 2012[c]	2011/12	July
Indonesia	DHS	2012	6–11[b]	31,297	May–Jul 2012	2011/12	July
Kenya	DHS	2003	6–13	10,451	Apr–Sep 2003	2003	Jan
Kenya	DHS	2008	6–13	11,006	Nov 2008–Feb 2009	2008	Jan
Rwanda	DHS	2005	7–12	10,850	Feb–Jul 2005	2004/05	Sep
Rwanda	DHS	2010	7–12	12,921	Sep 2010–Mar 2011	2009/10	Sep

[a] Unweighted. Individuals with nominal ages one year below to one year above the primary school age range.
[b] This is the *de facto* age range. The official *de jure* entry age in Indonesia is 7, however an analysis of datasets with exact birth date information shows that for at least the last 15 years, the vast majority of children have already entered school by the time they reach their 7th birthday.
[c] The survey actually consisted of four waves, but only the March and September waves were made available as public-use files including the wave information.

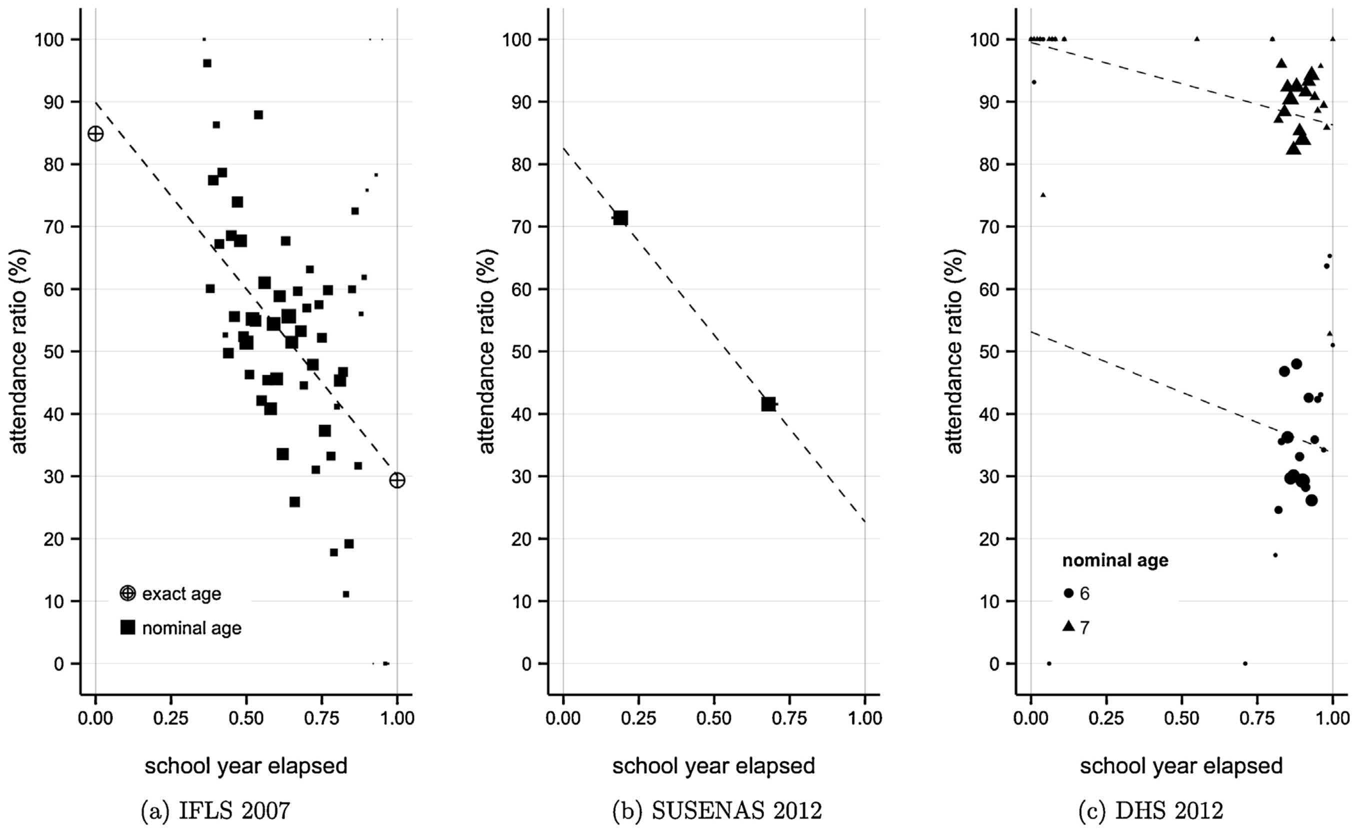

Fig. 2. Nominal ASAR at age 6 as a linear mix of true ASAR for ages 5 and 6.

Table 4
NAR (%) by age adjustment procedure for case study datasets.

Country	Survey	Year	NM	MB	IB	RD	RG	OD	Exact
Nigeria	MICS3	2007	54.6	59.2	58.5	57.0	60.1	64.1	
Nigeria	MICS4	2011	65.7	65.7	66.0	68.1	72.1	73.7	70.1[a]
Indonesia	IFLS	2007	90.1	97.1	95.9	94.6	96.0	96.9	95.9
Indonesia	SUSENAS	2012	91.6	91.6	96.1	95.1	95.5	97.7	
Indonesia	DHS[b]	2012	85.3	94.7	94.7	93.6	94.0	95.4	
Kenya	DHS	2003	78.7	78.7	80.1	81.2	82.9	84.1	
Kenya	DHS	2008	78.7	88.0	88.0	87.9	88.3	88.8	
Rwanda	DHS	2005	85.4	88.5	88.5	87.6	88.4	91.1	
Rwanda	DHS	2010	87.3	92.7	92.8	93.1	93.1	94.7	

[a] National Bureau of Statistics et al. (2013), based on birth months.
[b] Only the variable on current attendance contains usable information.
Note: For comparability and to compare the effect of the age adjustment procedures, all results shown are based on observed integer ages, even when more detailed age information is availably. These cases are noted.

Table 5
Primary ASAR at age 6 (%).

Survey	NM	MB	IB	RD	RG	OD
IFLS 2007	53.2	93.2	86.6	79.3	87.9	88.9
SUSENAS 2012	56.7	56.7	83.5	77.7	87.2	91.4
DHS 2012	36.0	87.7	87.7	81.9	83.9	87.6

complementary out-of-school ratio, this implies a relative difference of about 15 percent. Given the absolute numbers, this implies a difference in the order of magnitude of one million out-of-school children, or hundreds of millions of US dollars in the estimated cost of achieving universal primary education. Clearly such differences matter.

Turning to the question of timing, the example of Nigeria illustrates several points. First, there can be some arbitrariness involved in whether the binary adjustment is or is not applied, depending on how the exact date of school start and/or interviews are imputed, if these are known only to the nearest month. The Nigerian school year starts in 'September' according to international reports and databases; local experts confirm that the precise date varies by area or even school. Even a unique reference date may be challenging to determine for historical years to any greater than monthly precision. However, as Fig. 3 shows, the exact day matters for whether the median interview of the 2007 MICS, for example, occurred during the school year's first or the second half.

Second, still with reference to Fig. 3, consider assessing the change over time. For comparability, both waves should undergo the same adjustment; in particular, the more detailed age data

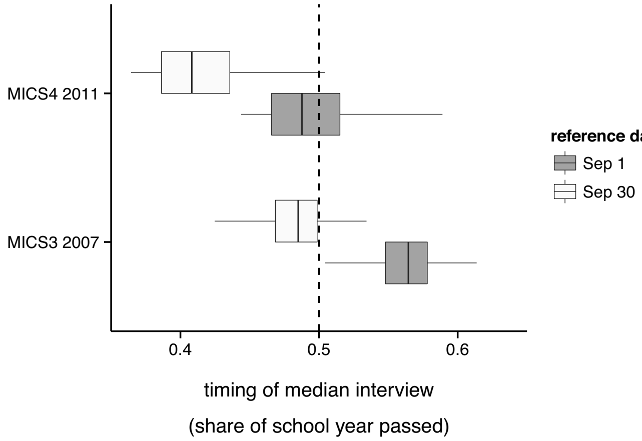

Fig. 3. Enumeration relative to school year, by exact reference date.
source: MICS.

available only for the latter wave cannot be exploited. More crucially, however, with surveys just either side of the threshold, as here, it is advisable to deliberately flout the threshold criterion and adjust either both or neither. In other words, it may be known that applying the binary adjustment procedure will *increase* the error in comparing different surveys. As mentioned, this forces an unpalatable choice between following a consistent process and obtaining consistent results.

Indeed, the difference between the individual and median binary adjustment may be framed similarly, only within a single survey. The IB adjustment makes the overall estimate more accurate at the expense of making individual observations potentially less comparable. The MB adjustment attempts to avoid this; unfortunately the same trade-off emerges at a different level of aggregation. For a given sub-national unit, the median interview may precede the threshold even if the overall national median exceeds it. Such a situation could well occur, as Fig. 4 shows: despite an impressive effort at synchronous enumeration across states, depending on the exact reference date (or if enumeration had shifted by mere days), some states had their median interview before the mid-year and others after. In this situation, applying MB at the national level amounts to knowingly *increasing* bias in the estimated NAR for some states. Unfortunately, applying MB at provincial level instead means the adjusted national NAR no longer equals the weighted average of the provincial values. The MB adjustment does not 'decompose'. The trade-off mentioned above re-appears: the threshold criterion applied to the national median makes state-level NARs more comparable, but the national NAR is then least accurate around mid-school-year; applying it to the state-level medians make state-level NARs less comparable, but – through partially off-setting errors – the aggregated national NAR most accurate around mid-year.

5.4. Kenya: implications for policy evaluation

The potential impact of unadjusted survey timing on policy conclusions is highlighted by a recent analysis of free primary education and school attendance in Sub-Saharan Africa, with specific reference to the cases of Kenya and Rwanda (Sandefur and Glassman, 2014). By juxtaposing administrative and survey data, the report argues that official enrolment statistics in these countries are not only highly unreliable, but systematically so, due to incentives faced by authorities to present primary school fee abolition as a policy success. However, the official narrative is seemingly disproved by independent household survey data on primary school attendance, which by contrast shows hardly any increase (Rwanda), or even none at all (Kenya) (c.f. Fig. 2 in Sandefur and Glassman (2014)).[12] This finding has been used to call into question the monitoring of development indicators altogether (Jerven, 2014).

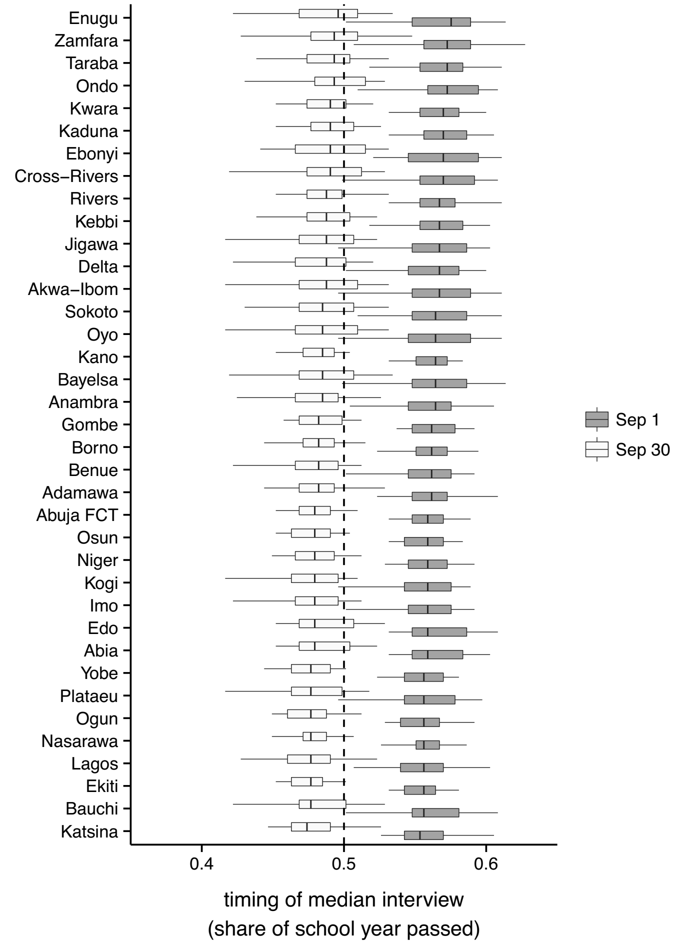

Fig. 4. Enumeration relative to school year, by exact reference date and state.
Source: MICS 2007.

Growing enrolments being contradicted by static attendance is indeed the impression created by the nominal-age NAR estimates in Fig. 5. Here, these nominal estimates were re-calculated from the DHS microdata, with unadjusted observed integer ages. That they match the values published in the DHS *StatCompiler* database (that also underlie the analysis in Sandefur and Glassman, 2014) is constructive proof that the published numbers are indeed entirely unadjusted for survey timing. The administrative enrolment figures shown in Fig. 6 are taken directly from Sandefur and Glassman (2014) as reported.[13]

Incidentally, both in Kenya and Rwanda, the more recent DHS was conducted later relative to the reference school year than the earlier survey (Fig. 6). Accordingly, the later NAR estimates

[12] Only the recent DHS figures are shown here. The additional surveys conducted by the Kenyan National Bureau of Statistics (KNBS) that Sandefur and Glassman show were also re-analysed for the this study, and attendance across providers matches the age-adjusted DHS trend well. While it may have been reasonable in the original report to refer to attendance at public schools only (as a comparison with Bold et al., 2011 confirms) given that an abolition of public school fees is the policy under evaluation, the disaggregation by provider is tangential to the issue at hand here.

[13] The original figures come from the Education Monitoring and Information System (EMIS) databases maintained by the respective Ministries of Education, via the World Bank's World Development Indicators (WDI) database. They are available as part of the replication dataset provided Sandefur and Glassman (2014).

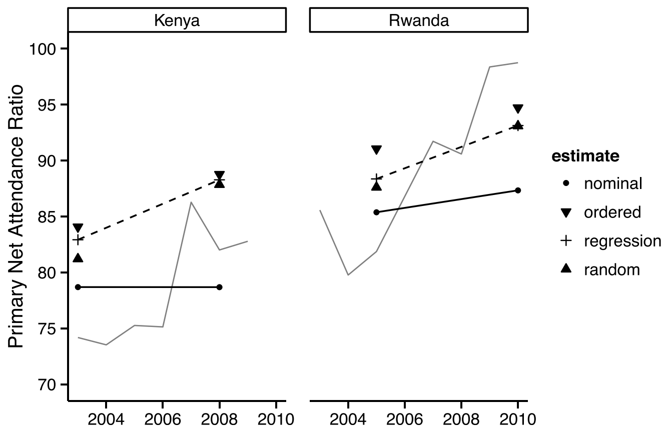

Fig. 5. Primary NAR by method of age adjustment for survey timing.
Source: own calculations from DHS microdata. Solid line: administrative enrolment figures from WDI database as reported in Sandefur and Glassman (2014).

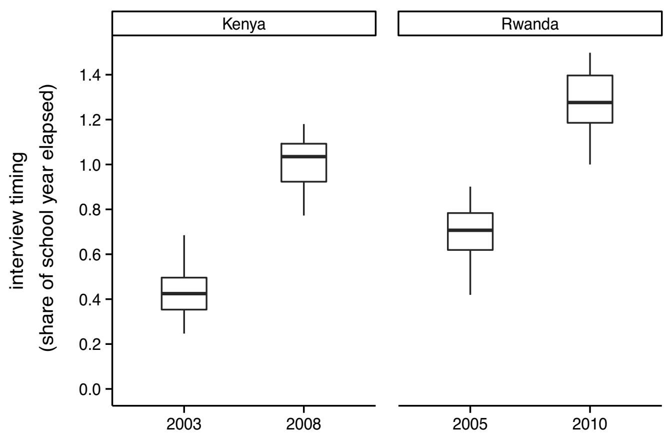

Fig. 6. Timing of enumeration relative to reference school year.
Source: DHS microdata.

accounting for survey timing are adjusted more strongly, altering the implied change over time. In fact, as Fig. 5 shows, for Kenya the slope based on adjusted household survey data matches the slope of the administrative indicators very well (though at a higher level overall).[14] In Rwanda, a gap between the increase in enrolment and attendance remains after adjusting for survey timing, but is strongly diminished and no longer requires deliberate misreporting as an explanation. So, far from showing that monitoring data is so poor as to not being worth collecting, these examples on the contrary demonstrate that there are 'zero-cost' gains available in terms of how existing data can be more fully exploited.

Note that Kenya and Rwanda were not arbitrarily selected, but because these countries exhibited the largest discrepancy between administrative and survey data. It is no surprise that the most extreme cases would be countries where survey timing contributes to this discrepancy. As it seems unlikely that the later household survey would have occurred later in the school year in all the countries considered by Sandefur and Glassman (2014), the results of their regression analysis may well hold regardless. However,

Table 6
Truncated primary NAR and ASAR at entry age, nominal.

Country	Year	Entry-age AR	Truncated NAR
Kenya	2003	40.7	84.4
Kenya	2008	28.8	86.7
Rwanda	2005	65.6	89.6
Rwanda	2010	65.6	92.5

Source: DHS microdata.

there is no way of knowing without re-estimating the model on age-adjusted data.

To demonstrate its potential utility, Table 6 shows that the simple truncated NAR does detect the overall increasing trend. Moreover, it shows that certainly in the case of Kenya, the sharply declining attendance of the nominal entry-age cohort contrasted with the rising truncated NAR could have served as a warning signal that measurement issues and/or age dynamics might be distorting the face-value results.

6. Discussion and conclusion

An effective adjustment for the survey timing effect continues to be relevant. Recording integer ages remains a common practice; in any case, existing datasets with integer ages are here to stay, and analysing indicator time series requires comparable 'lowest common denominator' procedures for all waves. Moreover, in low-education environments, respondent reports of birth dates are often incomplete and possibly unreliable; estimates not reliant on these responses are necessary to fill gaps and as a validity check.

Given integer age data, there is a demonstrable potential of unadjusted survey timing – or of misunderstood partial corrections – to result in an estimation error that, at double-digit percentage points, is just as large as the trends or policy impacts we actually wish to study in international educational development. Uncorrected, an error of such size can evidently distort research results and their interpretation, or lead to the unwitting attribution of spurious credit or blame in the policy arena. As shown above, this potential is actualised too often, and the possible distortion too large, to continue with the status quo. Currently, the burden effectively rests on individual researchers to realise that in order to avoid a problem that goes unmentioned in the reference works they may turn to for authoritative guidance, they must wade through technical reports and/or the processing syntax of data providers. Even if they do, they would currently encounter recommendations that are at best partial and conflicting, and at worst incorrect.

Fortunately, careful analysis shows that we actually do possess a good analytic handle on both the direction and magnitude of the error, and that this promise is borne out in empirical application. Individual researchers analysing only a few countries can investigate the appropriateness of the intensely data-driven regression-based point estimate presented here, and consider the implications of the bounds. For agencies, reporting bounds and intervals may be at odds with the communication constraints imposed on reports that target the general development community and policy makers, an audience that is generally believed to strongly prefer point estimates. Actually, a coherent point estimate is available in form of the mid-point of the interval between the bounds. In light of the manifest weaknesses of current procedures, a perception that they might be easier to explain than the procedures suggested here is not a good reason to maintain the status quo. In any case, current practice is not in fact perceived as transparent or interpreted correctly, despite being simple or perhaps even simplistic.

[14] Given that enrolment indicators more typically exceed estimates of actual attendance, candidate explanations might be that the population in the denominator of the administrative enrolment ratios is overestimated, or that the administrative figures do not capture certain types of schools. In any case, the difference in interpretation between the nominal NAR estimates and the age-adjusted ones is about the absence or presence of change over time, so the level shift is not investigated further here.

Schooling data and associated indicators suffer from numerous potential errors, biases, and distortions, independently of the survey timing effect, of course. Some of these may very well be larger in magnitude on occasion. However, the examples shown here, and the presentation in terms of absolute changes in NAR, are actually conservative. In terms of relative changes to the complementary *out-of-school rate*, the effect of adjusting for survey timing is even greater than the errors shown in this study, further highlighting the need for additional methodological research.

Acknowledgements

I thank the two anonymous referees as well as Stephanie Bengtsson for their assiduous reading of a demanding manuscript and their invaluable critique.

Appendix: (Mis-)specifying a survey-timing adjustment for age-appropriate enrolment/attendance

In the following, U, T, O^1, O^{2^+} stand for the true under-age, on time, 1 year over-age, and 2 or more years over-age enrolment at the beginning of the school year, and the variants with p subscripts stand for the corresponding values observed after a fraction p of the school year has elapsed. The attempted adjustment specified in EPDC, 2009 can be summarized in matrix form thus:

$$\begin{pmatrix} U \\ T \\ O^1 \\ O^{2^+} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & p & 0 & 0 \\ 0 & 1-p & p & 0 \\ 0 & 0 & 1-p & p/2 \\ 0 & 0 & 0 & 1-p/2 \end{pmatrix}}_{M} \times \begin{pmatrix} U_p \\ T_p \\ O^1_p \\ O^{2^+}_p \end{pmatrix}$$

Multiplying both sides by the inverse matrix M^{-1} uncovers the implied relationship between the observations part-way through the school year and the true values at the beginning of the school year:

$$\begin{pmatrix} U_p \\ T_p \\ O^1_p \\ O^{2^+}_p \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & p/p-1 & p^2/(p-1)^2 & p^3/(p-2)(p-1)^2 \\ 0 & 1/1-p & -p/(p-1)^2 & -p^2/(p-2)(p-1)^2 \\ 0 & 0 & 1/1-p & -p/(p-2)(p-1) \\ 0 & 0 & 0 & -2/p-2 \end{pmatrix}}_{M^{-1}} \times \begin{pmatrix} U \\ T \\ O^1 \\ O^{2^+} \end{pmatrix}$$

This implied relationship is logically impossible, *whatever the true school flows*, because – contrary to what is claimed here – over-age students cannot turn into under-age students over the course of the year. This contradiction has nothing to do with the simplifying assumption that half of those 2 or more years over-age are exactly 2 years over-age, and remains true whatever is assumed in that respect. Formally, M^{-1} contains non-zero entries in positions that should actually be structural zeros.

Indeed, the most general specification of how the observations throughout the school year derive from the starting conditions is

$$\begin{pmatrix} U_{p_i} \\ T_{p_i} \\ O^1_{p_i} \\ O^{2^+}_{p_i} \end{pmatrix} = \begin{pmatrix} q_i^U & 0 & 0 & 0 \\ p_i^U & q_i^T & 0 & 0 \\ 0 & p_i^T & q_i^{O^1} & 0 \\ 0 & 0 & p_i^{O^1} & 1 \end{pmatrix} \times \begin{pmatrix} U \\ T \\ O^1 \\ O^{2^+} \end{pmatrix},$$

where the birth month distribution may differ between categories, drop-out is possible, and the p_i for different months need not be linearly related. It is evident that no instance of this can lead to M^{-1} above implied by the adjustment in (EPDC, 2009).

With the correct specification of uniformly random birth months, i.e. $q_i = 1 - p_i$ and $p_i = i/12$, $i = 1, 2, \ldots, 12$ for all categories, the adjustment becomes

$$\begin{pmatrix} U \\ T \\ O^1 \\ O^{2^+} \end{pmatrix} = \begin{pmatrix} 1/1-p & 0 & 0 & 0 \\ -p/(p-1)^2 & 1/1-p & 0 & 0 \\ -p^2/(p-1)^3 & -p/(p-1)^2 & 1/1-p & 0 \\ p^3/(p-1)^3 & p^2/(p-1)^2 & p/p-1 & 1 \end{pmatrix} \times \begin{pmatrix} U_p \\ T_p \\ O^1_p \\ O^{2^+}_p \end{pmatrix}.$$

Note that there is no need to make any specific assumptions about the distribution within the O^{2^+} category. The fact that the 'back-cast' value for over-age students is related not just to on-time students, but also under-aged ones, despite the fact that individuals can experience at most one birthday during the school year, does not pose the same kind of logical contradiction as the incorrect specification of M^{-1}. It reflects the cascade of upward transitions and that determining the share of children that has been added to a category since the beginning of the school year depends on knowing the initial size of the level below, which in turn depends on the levels even further below.

References

Barakat, B., Durham, R., Rodrigues Guimarães, C., 2013. Age compositional adjustments for educational participation indicators. Population 68 (4), 607–626.

bildungsserver.de, 2014, September. Zur Stichtagsregelung in den Bundesländern, http://www.bildungsserver.de/innovationsportal/bildungplus.html?artid=846 (accessed 20.11.14).

Bold, T., Kimenyi, M., Mwabu, G., Sandefur, J., 2011. Why did abolishing fees not increase public school enrollment in Kenya? Center for Global Development Working Paper Series 271. .

Boyden, J., 2014. Young Lives: An International Study of Childhood Poverty: Rounds 1–3 Constructed Files, 2002–2009. UK Data Archive.

Carr-Hill, R., 2012. Finding and then counting out-of-school children. Compare: J. Compar. Int. Educ. 42 (2), 187–212.

Case, A., Paxson, C., Ableidinger, J., 2004. Orphans in Africa: parental death, poverty, and school enrollment. Demography 41 (3), 483–508., http://www.jstor.org/stable/1515189.

EDOREN, 2014. Education Data in Nigerian National Household Survey Data: Review of Survey Reports and Evidence from Microdata. Education Data, Research and Evaluation in Nigeria (EDOREN), Abuja, Nigeria.

Education Policy and Data Center (EPDC), 2009. 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.

EPDC, 2009. Pupil performance and age: a study of promotion, repetition, and dropout rates among pupils in four age groups in 35 developing countries, Working Paper 09-02, Education Policy and Data Center.

Fox, L., Santibaez, L., Nguyen, V., Andr, P., 2012. Education Reform in Mozambique: Lessons and Challenges. World Bank Publications.

Glewwe, P., 2000. Education. In: Grosh, M., Glewwe, P. (Eds.), Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study, vol. 3. World Bank, Washington, DC.

Glewwe, P., Kremer, M., 2006. Schools, Teachers, and Education Outcomes in Developing Countries. Vol. 2 of Handbook of the Economics of Education. Elsevier, pp. 945–1017. (Chapter 16), http://www.sciencedirect.com/science/article/pii/S1574069206020162.

ICF International, 2013. Tabulation Plan for DHS Final Report. Demographic and Health Surveys Methodology. USAID.
IPUMS, 2014. School Attendance: Questionnaire Text, In: https://international.ipums.org/international-action/variables/SCHOOL/#id2010a (accessed 20.11.14).
Jerven, M., 2014. Benefits and Costs of the Data for Development Targets for the Post-2015 Development Agenda. Data for Development Assessment Paper. Copenhagen Consensus Center.
National Bureau of Statistics, UNICEF, UNFPA, 2013. Nigeria Multiple Indicator Cluster Survey 2011: Main Report.
National Population Commission (Nigeria), RTI International, 2011. Nigeria DHS EdData Survey 2010: Education Data for Decision-making. National Population Commission and RTI International, Washington, DC.
Omoeva, C., Sylla, B., Hatch, R., Gale, C., 2013. Out Of School Children: Data Challenges in Measuring Access to Education. fhi360. Education Policy and Data Center (EPDC), Washington, DC.
Orazem, P.F., King, E.M., 2007. Schooling in developing countries: the roles of supply, demand and government policy. Handb. Dev. Econ. 4, 3475–3559.
Porta, E., Arcia, G., Macdonald, K., Radyakin, S., Lokshin, M., 2011. Assessing Sector Performance and Inequality in Education. World Bank.
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.
RAND Corporation, 2014. Indonesia Family Life Survey (IFLS), In: http://www.rand.org/labor/FLS/IFLS.html (accessed 20.11.4).
Sandefur, J., Glassman, A., 2014. The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics, http://dx.doi.org/10.7910/DVN/26712.
Sandefur, J., Glassman, A.L., 2014. The Political Economy of Bad Data: Evidence from African Survey & Administrative Statistics Available at SSRN 2466028.
Smiley, A., Omoeva, C., Sylla, B., Chaluda, A., 2012. Orphans And Vulnerable Children: Trends in School Access and Experiences in Eastern and Southern Africa. fhi360 Education Policy and Data Center (EPDC), Washington, DC.
Statistics Indonesia (BPS), 2012. National Socio-economic Survey (SUSENAS), In: http://microdata.bps.go.id/mikrodata/index.php/catalog/SUSENAS (accessed 20.11.14).
Strauss, J., Witoelar, F., Sikoki, B., M. W.A., 2009. The fourth wave of the Indonesian Family Life Survey (IFLS4): Overview and field report. WR-675/1-NIA/NICHD.
UIS, 2004. Guide to the Analysis and Use of Household Survey and Census Education Data. UNESCO Institute for Statistics.
UIS, 2005. Children Out of School: Measuring Exclusion From Primary Education. UNESCO Institute for Statistics.
UIS, 2010. Measuring educational participation: analysis of data quality and methodology based on ten studies. Technical Paper 4. UNESCO Institute for Statistics.
UIS, 2012. International Standard Classification of Education ISCED 2011. UNESCO Institute for Statistics.
UNESCO Division of Statistics, 1997. Primary and Secondary Education: Age-Specific Enrolment Ratios by Gender 1960/61-1995/96. UNESCO.
UNICEF, 2012. Nigeria Country Study. Global Initiative on Out-of-School Children. UNICEF.
UNICEF, 2014. South Asia Regional Study. Global Initiative on Out-of-School Children. UNICEF.
UNICEF and UNESCO Institute for Statistics (UIS), 2015. Global Out-of-School Children Initiative Operational Manual. UNICEF.
Wils, A., 2004. Late entrants leave school earlier: evidence from Mozambique. Int. Rev. Educ. 50 (1), 17–37.
World Bank, Instituto Nacional de Estatistica Moçambique, 2008. Mozambique Education Outcomes National Panel Survey (NPS). Technical Document, In: http://microdata.worldbank.org/index.php/catalog/999/download/20596.