Annual Meeting, Population Association of America

New Orleans, LA, April 11-13, 2013

A New Approach to Indirect Estimation of Child Mortality: Application to Malawi

Agbessi Amouzou[1] and Kenneth Hill[1]

[1] Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe street, Baltimore, MD 21205

Agbessi Amouzou: aamouzou@jhsph.edu; Kenneth Hill: khill@jhsph.edu

Abstract:

Information from women's summary birth history consisting of only two questions - one on number of children ever borne alive and the other on children dead – is used traditionally for indirect child mortality estimation. However the estimation is based on several assumptions about fertility and mortality patterns and recent rates computed are biased. We propose and apply an innovative approach based on imputation of an existing full birth history onto summary birth history data using the summary birth history information. The resulting imputed full birth history is used to calculate child mortality rates using standard life table procedures. We applied the approach to the Malawi 2008 Population Census and the 2004 and 2010 demographic and health survey datasets. Preliminary results show very good promise of the approach with the imputed child mortality rates falling within 20% of less of the rates directly computed from the 2010 DHS survey.

Primary session: **411: Maternal, Infant, and Child Health and Mortality**

Secondary session: **413: methodological issues in health and mortality**

Introduction

At the Millennium Summit in 2000, the United Nations and International Organizations set forth the Millennium Development Goal 4 to reduce mortality among children under-five by two-third between 1990 and 2015 in developing countries. Monitoring this objective requires frequent data to estimate under-five mortality level at national and subnational level with good enough precision that would allow detecting statistically significant trends. Although most low and middle-income countries are participating in demographic and health survey (DHS) programs that carry out national surveys every four or five years, these data do not provide frequent and recent enough mortality estimates to effectively monitor the decline in mortality and adjust child survival programs consequently. Given the slow pace of decline in mortality observed especially in poor countries, "real-time" monitoring of mortality is advocated as a strategy for tracking country's progress and making country more accountable. The "real-time" mortality monitoring implies measurement of child mortality on short term periods, no longer than annual. The need for real-time child mortality monitoring is further challenged by the new commission on Accountability for Women Children's Health that has been established by the United Nations to monitor progress in the MDGs.

This strategy can only be implemented if good quality mortality data are available every year for computation of annual under-five mortality rates. To dates, countries have estimated childhood mortality directly by collecting survey data on full birth history from women aged 15-49, typically implemented through DHSs or Multiple Indicators Cluster Surveys (MICS). Data collection on full birth history in low and middle income countries requires tedious and careful training and preparation and close supervision in the field to ensure data quality needed for accurate mortality estimation. The burden of activities to carry out during those surveys usually requires that sample size of household be limited to a minimum level, generally lower than the level that is required to generate annual mortality estimates with acceptable precision. In addition because the surveys usually include many other modules, they are expensive to carry out every year. Aside from a full birth history, childhood mortality can be estimated indirectly using summary birth history data from women age 15-49 and applying the traditional methodology pioneered by William Brass. A summary birth history consists of two questions posed to women aged 15-49 on their number of children ever borne and their number of children who have died among those ever borne. Additional information is obtained on the exposure of children to death by collecting the age of the women, the time since her first birth or the time since first marriage. These questions have been used to compute mortality indirectly based on several assumptions about the patterns of mortality and fertility.

A summary birth history data collection can be carried out fairly easily than a full birth history and requires less intensive training and supervision. It can be included in national annual surveys designed for other purposes allowing possibility to compute child mortality estimates indirectly. It can also be – and has been – incorporated in population census questionnaires. This is particularly useful because it provides possibility to estimate mortality at small geographic area level. However, traditional indirect mortality estimates do not provide reliable recent estimates and are heavily dependent on the choice of particular mortality model. Mortality estimates derived are usually based on five-year age group of mothers and the most recent mortality estimate obtained is based on information from women 15-19.

This estimate is known to be highly over estimated because of the selection bias among women reporting experience of birth in that young age group. Typically, these mothers aged 15-19 generally come from lower socio-economic group and the births that they report are also mostly first births. To reduce this selection, other exposure variables such time since first birth or time since first union or marriage are used instead of the age of the mother. However these variables are commonly more affected by quality issues than age and the age-based mortality estimation is recommended. Indirect mortality estimates based on summary birth history also do not take into account the distribution of births and deaths and rely only on the proportion dead among children ever borne.

Due to these weaknesses in the indirect methods, countries have preferred the direct estimation method, at the expense of timeliness and cost. Proper and effective real-time mortality monitoring requires innovative approaches that take advantages of the simplicity of data collection on summary birth histories and are able to produce timely mortality estimates that countries can use to assess effectiveness of their child survival programs. The simplicity of data collection on summary birth history means that survey sample size can be increased sufficiently to generate estimates that are as recent as annual.

We propose and apply in the context of Malawi an innovative approach to childhood mortality estimation that uses data on women's summary birth history and borrows full birth history information from existing surveys to calculate child mortality. The main idea is the use of full birth history data from an existing survey to impute on SBH recent data based on of the summary birth history information. We describe below the principles of the method.

Basis of birth history imputation

Although women's summary birth history is traditionally used for indirect mortality calculation, it cannot provide information about child mortality in a specific (short) recent time because the children reported by women are spread over a number of previous years, and are exposed to potentially time-varying mortality risks. The only exception would be children born to women who report their first birth as occurring in the 12 months before the interview, but this sample will be entirely first births and thus non-representative of all children.

The birth history imputation approach uses data on summary birth history and borrows full birth history information from an existing data. Individual women from the two datasets are matched based on their number of children ever born alive (CEB), the number of children who died among those born alive (CD) and age of the mother (or time since first birth if available in both datasets). Typically the approach matches each woman who responded to the two summary birth history questions (Woman "A") to a woman with similar CEB, CD and age group (five-year age group) for whom a full birth history was collected in another survey – for example, a DHS survey (Woman "B"). The full birth history details of Woman "B" are then assigned to Woman "A" and used to calculate under-five mortality. To reduce noise, each summary birth history would have multiple randomly-selected full histories with the same matching characteristics attached to it. Some relatively small proportion of women may not have an exact match on all three variables from the two datasets, in which case the match is made on CEB and CD only. Because the summary birth history data would reflect changes in overall mortality since the full birth history was collected (since the distribution of women by children ever born and children dead would change) the mortality estimates derived from this method should at least partially capture recent mortality trends. The approach relies mainly on the assumption of average time invariance in the distribution of births and deaths in the full birth history given the women CEB, CD and age. In other words the time distribution of births and deaths in a full birth history given women's CEB, CD and age, does not on average depend on time of observation. If F1 and D1 are the distribution of births and deaths at time t1 given CEB, CD and age; and F2 and D2 are distribution of births and deaths given CEB, CD and age at time t2, the assumption translates to:

$$\boldsymbol{F_{1|CEB,CD,Age} = F_{2|CEB,CD,Age}} \quad \textbf{and} \quad \boldsymbol{D_{1|CEB,CD,Age} = D_{2|CEB,CD,Age}}$$

The imputed full birth history derived is used to calculate childhood mortality using life table procedures similar to a true full birth history data. The approach has several potential advantages compared to the traditional indirect method based on summary. First, it relies only on the assumption of no temporal systematic variation in the distribution of births and deaths given CEB, CD and age of the woman. Unlike the traditional indirect estimation methods, it makes no assumption about fertility or mortality patterns and is less affected by fertility trends. Second, because it uses information from all women 15-49 years recent mortality estimates are not affected by the selection bias observed in traditional indirect estimation for which estimates based on young mothers are usually biased. Finally and most important advantage is its applicability to large datasets such as population census data, thus allowing estimation recent or annual child mortality trends at small geographic area level. Summary birth history can be incorporated in national surveys designed for other purposes (e.g. SMART survey, Welfare Monitoring Surveys, Integrated Household Survey) and data can be used to estimate annual mortality using the imputation approach. We present in the next section data and method used to apply this approach in the context of Malawi and present the results.

Data

We apply the birth history imputation approach using data from the 2008 Malawi Population and Housing Census (PHC) and the 2004 and 2010 Demographic and Health Surveys. The 2008 Malawi PHC included a module on summary birth history that collected information on number of children ever borne and number of children surviving from women aged 12 and above. We use a 10% sample of the PHC data, corresponding to 309,851 women 15-49 and took another 10% random sample of these women. Thus the total sample of women age 15-49 from the PHC data was 30,985. The 2004 and 2010 DHSs included full birth history data from women aged 15-49. The 2004 collected data from 13,664 households and 11,698 individual women aged 15-49. The 2010 survey collected data from 24,825 households and 23,020 individual women. SBH data from the PHC was used and two separate imputed full birth history was generated using separately the 2004 and 2010 DHSs.

Analytical method

In each dataset, we categorized mothers by CEB, CD and five-year age group and within each category women were sorted based on a random function. Women in the PHC data were then matched with women from the DHS data based on categories defined by CEB, CD and age. Because the PHC data had more mothers than the DHS data, we ensured that the DHS data is expanded within each category in order to obtain the same number of records within each category between the two datasets before the merge is conducted. In doing so, a woman from the DHS category defined by CEB, CD and age group may be matched to multiple women of the same category in the PHC data. Multiple imputations are conducted by repeating the process ten times. Two imputed full birth history data are generated, one based on the 2004 DHS and the other based on the 2010 DHS. After adjusting the distributing of births based on the data of interview of the PHC, these data were then used to compute annual neonatal, infant, and under-five mortality rates for the past ten years preceding the PHC data collection, 1998-2007. We compute the same direct mortality rates using the DHS 2010 datasets and compared the results with imputed mortality rates. We computed the 95% confidence intervals of the direct mortality rates computed from the 2010 DHS data using jackknife resampling method.

We tested the main assumption by computing and comparing between 2004 and 2010 the average time since birth (that is age of child if child is alive or age the child would have had if alive for children dead) for all births in the same categories defined by CEB, CD, age group and birth order. The main assumption states that the distribution of births and deaths for the same women with same CEB, CD and age is on average time invariant. This implies that the average time since birth for all births for the same category of women characterized by same CEB, CD and age would be similar between 2004 and 2010. To test the time invariance in the distribution of births by mothers' CEB, CD and age group, we pooled the 2004 and 2010 DHS data and regress the average time since birth onto time (2004 or 2010) and a categorical variable (CODE) created by concatenating mothers' CEB, CD, age group and birth order. The regression model is

$$Average\ time\ since\ birth = \beta_0 + \beta_1 time + \beta_2 CODE + \varepsilon$$

time is 0 if DHS 2004 survey and 1 if DHS 2010 survey. The regression coefficient β_1 expresses the difference in the average time since birth between 2004 and 2010 adjusting for each women category defined by the variable CODE. We test whether the coefficient β_1 is significantly different from zero. The assumption is supported if β_1 is not statistically different from zero.

We proceed similarly to test the assumption for the distribution of deaths using the average age at death and the average time since death within each category defined by CEB, CD and age group. For these cases we no longer take into account the birth order of the child.

Results

Test of main assumption

Table 1 below shows the regression coefficient of the average time since birth, the average time since death and the average age at death on the survey year. Its compares the average of these dependent variable between 2004 and 2010, controlling for categories of women defined by CEB, CD, age and birth order. The regression indicates that all three coefficients are not statistically different from zero suggesting that the averages are statistically the same for 2004 and the 2010 survey. This results support the main assumption of time invariance in the time distribution of birth and death of a full birth history given CEB, CD, and age of the women.

Table 1: Linear regression coefficient of average time since birth, average time since death and average age at death on survey year (2004 versus 2010)

Dependent variable	Survey year (2004 versus 2010)				
	Coefficient	P-value	95% CI		N
Average time since birth*	-1.04	0.7040	-6.41	4.33	4377
Average time since death[$]	5.86	0.2700	-4.55	16.27	521
Average age at death[$]	-1.63	0.2620	-4.49	1.22	521

* Average within women categories defined by CEB, CD, age group and birth order
[$] Average within women categories defined by CEB, CD, age group

Preliminary Imputation results

Neonatal mortality

Figures 1a presents the imputed neonatal mortality rates using summary birth history data from the 2008 Malawi PHC and borrowing full birth history from the 2004 and 2010 DHS data. The trends are also compared to the direct neonatal rates computed from the 2010 DHS data. Overall the imputed neonatal mortality rates are fairly close to the direct rates, especially for the period 1999 and 2002. Although the imputed neonatal rates are the same in 2003 for the two datasets, they slightly under-estimate the rates compared to the 2003 direct rate. There appear more discrepancies toward recent period. Figure 1b show the ratio of each imputed rate to the corresponding direct rate from the DHS 2010 datasets. It indicates two main results. First it shows that the imputed neonatal rates are overall within 20% of the direct neonatal rates, except for the year 2006. Second, it shows that the full birth history data used for the imputation doesn't really matter. The imputed rates using either the 2004 DHS or the 2010 DHS are very close, especially between 1998 and 2003.

Infant mortality rates

Figure 2a presents the imputed infant mortality rates using the 2004 DHS and the 2010 DHS, and the direct estimates from the 2010 DHS. The imputed infant mortality rates are also very close to each other and close to the direct estimates from the 2010 DHS. The imputed rates using the 2004 DHS is particularly close to the direst estimate from the 2010 DHS and reflect almost perfectly the trends observed in the 2010 DHS. The results are confirmed in figure 2b showing the ratios of the imputed rates to the direct rates. Almost all imputed rates using the 2004 DHS fall within 10% of the direct infant mortality rates from the 2010 DHS and the imputed rates using the 2010 DHS fall within 20% of the direct rates.

Under-five mortality rates

Figure 3a and 3a show the imputed mortality rates and the ratio of the imputed rates to the direct 2010 DHS rates. Similar to the infant mortality rates, the imputed rates are very close to each other and very close to the direct estimate. The imputed rate observed in 2003 appears to depart slightly from the general trends and require further checking. Except for the 2003 under-five mortality rates, all imputed under-five mortality rate fall within 20% of the direct estimates.

Figure 1a: Imputed neonatal mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 and 2010 DHS, and direct neonatal mortality estimates from the 2010 DHS, 1998-2007

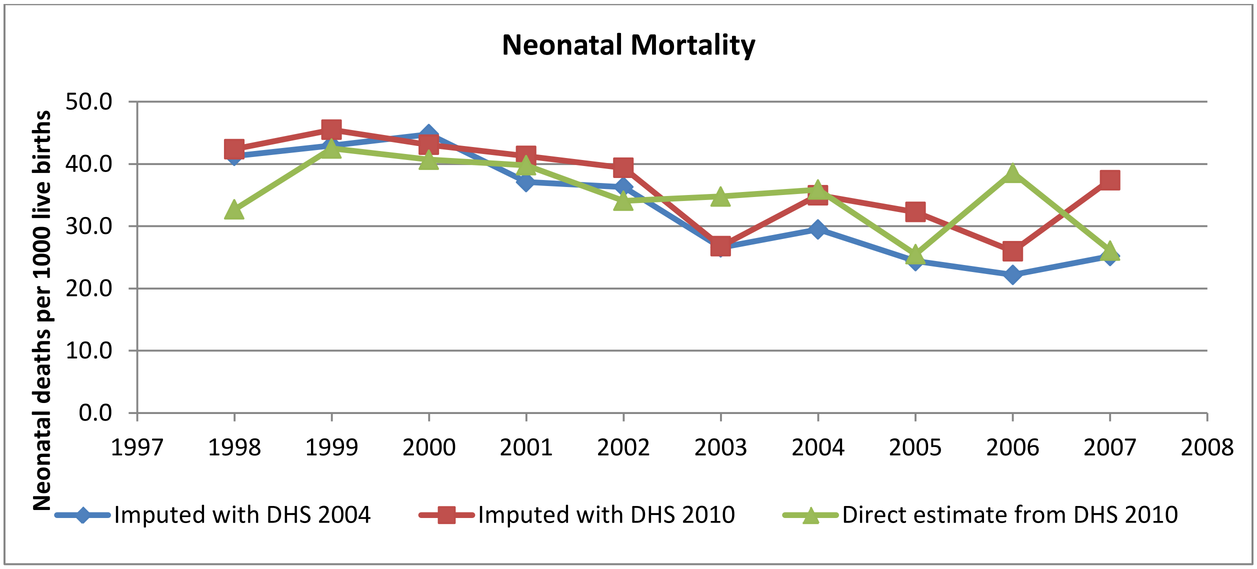


Figure 1b: Ratios of imputed neonatal mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 and 2010 DHS, to direct neonatal mortality estimates from the 2010 DHS

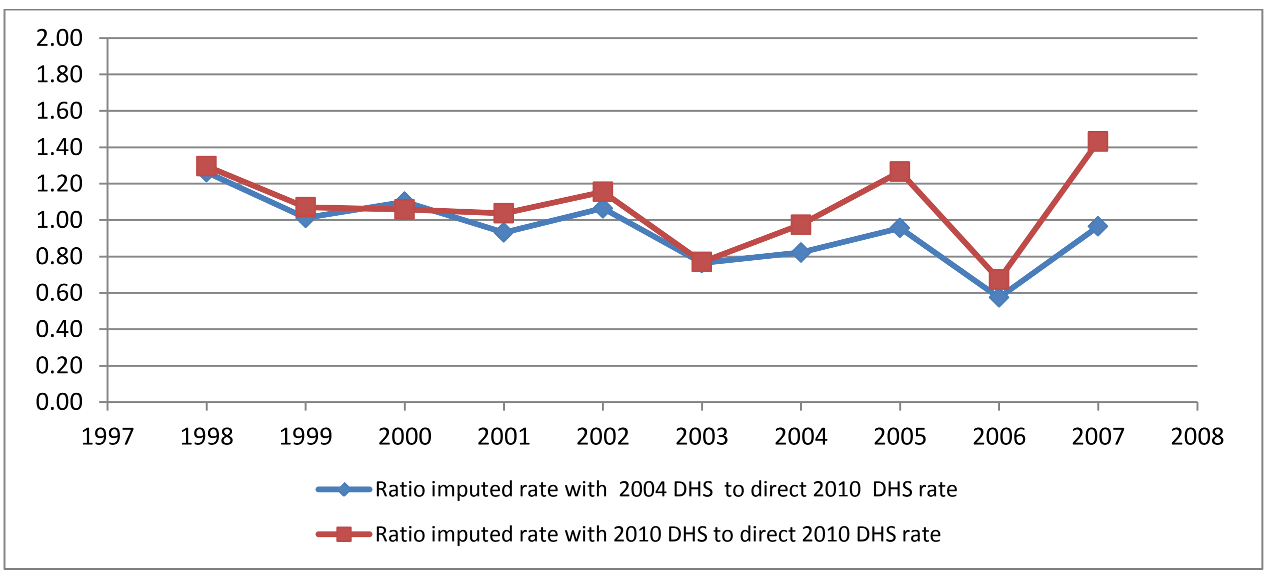


Figure 2a: Imputed infant mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 ad 2010 DHS, and direct infant mortality estimates from the 2010 DHS, 1998-2007

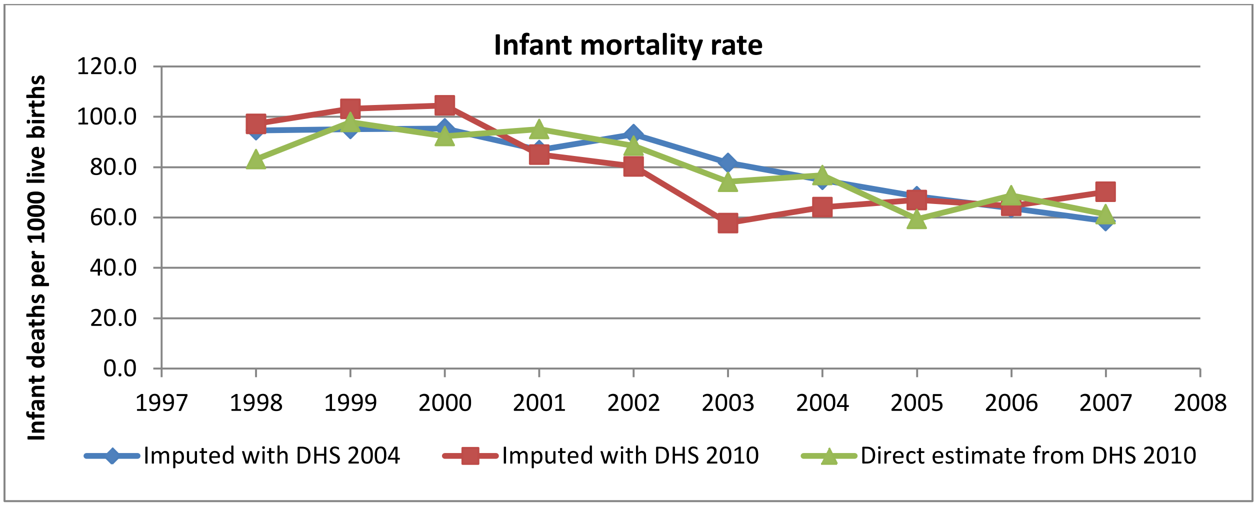


Figure 2b: Ratio of imputed infant mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 ad 2010 DHS, to direct infant mortality estimates from the 2010 DHS

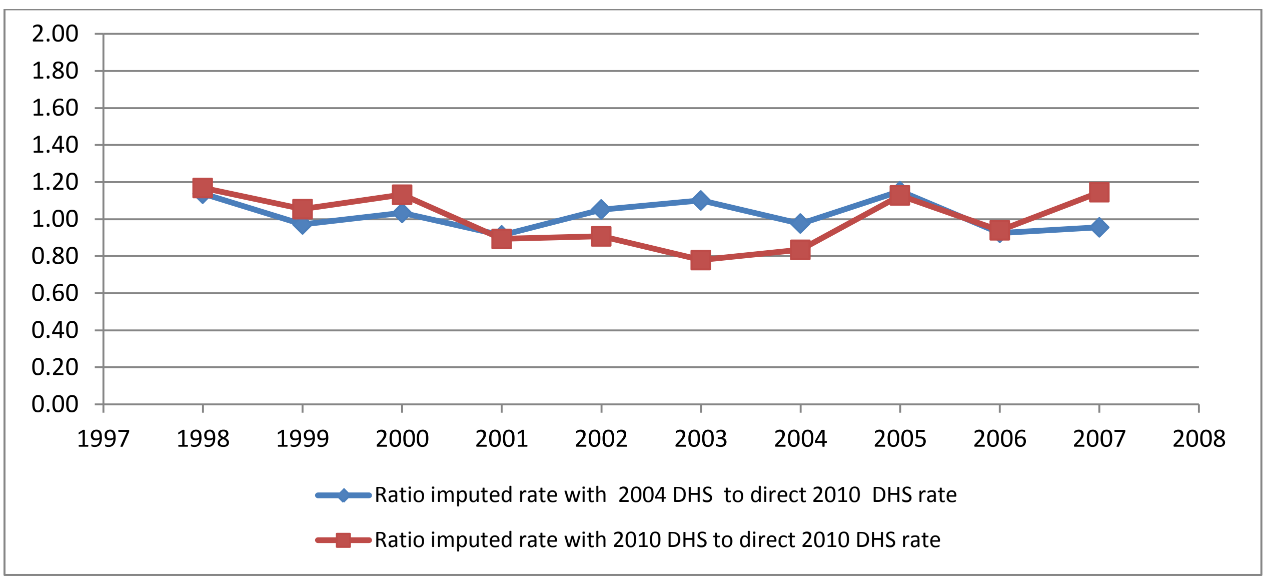


Figure 3a: Imputed under-five mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 and 2010 DHS, and direct under-five mortality estimates from the 2010 DHS, 1998-2007

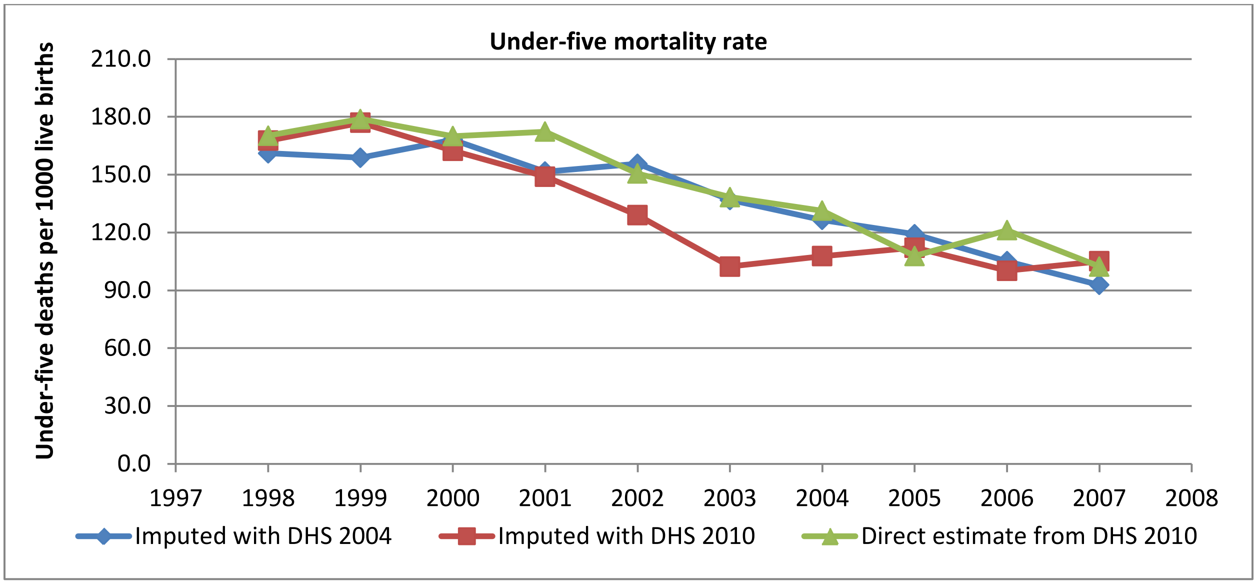
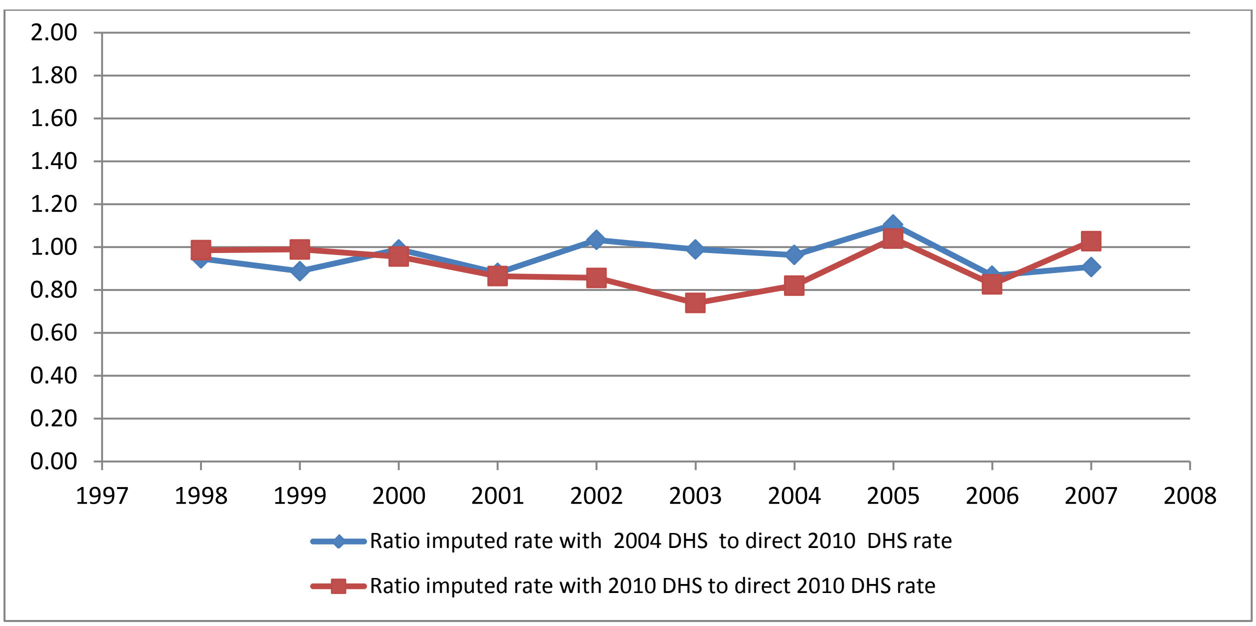


Figure 3b: ratios of imputed under-five mortality rates using summary birth history data from the Malawi 2008 PHC and borrowing full birth history data from the 2004 and 2010 DHS, to direct under-five mortality estimates from the 2010 DHS



Discussion

The preliminary results of the imputation approach using either women's full birth history data from the 2004 DHS or the 2010 DHS data suggests that the approach is very promising. Annual neonatal, infant and under-five mortality rates computed based on the imputed data are very close to the corresponding rates directly computed from the 2010 DHS, falling for almost all the rates within 20% of the direct rates. We are currently assessing further the approach and calculating the 95% confidence intervals to determine whether the imputed rates fall within the confidence interval. However, given the mortality rates presented refer to annual periods, fluctuations due to smaller samples size over these short periods could have affected the rates. The approach nevertheless captures fairly well both the trends and the levels of mortality rates. In addition the mortality rates computed using either the 2004 DHS or the 2010 DHS data for the imputed are also very close, suggesting that the full birth history data used to carry out the imputation does not matter.

Success of this approach has several advantages for countries, including the possibility to include summary birth history information in surveys conducted annually by national statistics offices and data used to compute child mortality. The approach represents an effective approach for monitoring child mortality on short term intervals such as annual periods given the opportunity to piggy-back on other surveys or conduct simple stand-alone surveys. In addition, given population censuses include data on summary birth history, countries can apply this imputation approach to estimate mortality in small geographic areas such as districts or sub-districts and on short reference periods. Data from national surveys such as DHS are currently limited to producing mortality at regional level for most countries and on periods generally as large as ten years preceding the survey. This has made it difficult for countries to obtain district level data on mortality to assess whether child survival programs implemented at these levels are producing the expected effects. With the imputation approach, simple surveys can be conducted at these geographic levels to collect data on women's summary birth history, and apply the approach to generate levels and trends in child mortality rates. We are conducting further analysis to confirm the applicability of the approach at sub-national levels.