Assessment of Sudan Households Health Survey Data, 2010

Awatif El Awad Musa (Corresponding author), and Hagir Osman El Jack

Al Neelain University, Faculty of Technology of Mathematical Sciences and Statistics

Mob: +249115059802, Fax: +249183789055

Email: awatif_musa@hotmail.com

Abstract

Most of the measures produced from Household Health Surveys depend to a critical degree on reports of ages and dates. The objectives of this paper are: to assess three aspects of SHHS, 2010 data quality, namely: women's age data, the children year birth, age in months, and anthropometric data. The data used is the raw data set file of the Sudan Household Health Survey 2010 (SHHS, 2010). Myers' Index is calculated to measure the age heaping of women, the birth year ratio is calculated to examine the displacement of children's age and final decimal digits of weight and height readings ending in.0 and .5 are also used for anthropometric data. Myers' Index shows that there is heaping in women ages at '0'and '5. Age heaping is higher in age of rural women than urban ones. Women educational level tends to affect age heaping, (lowest in age of women possess secondary and higher educational level and highest in age of illiterate women). Year birth ratio shows underestimation in births in the second and fourth years and overestimation in third year preceding the survey. Anthropometric data show high heaping in height measures slight in weight ones.

Keywords: data quality, displacement, heaping, Myers Index, anthropometric, height, weight.

1. Introduction

Household health surveys in Sub-Saharan Africa have the high levels of age displacement, Sudan is no exception, and however, with its long experience in collecting health data through successive national household health surveys; noticeable improvement in data quality in recent household health surveys is evident [2].

Completeness and accuracy of reporting dates and ages are critically important in household health surveys because the eligibility for inclusion in the survey of women age 15-49, as well as most surveys of men and special surveys, depends on the age given in the household survey. Eligibility of children for questions on health also depends on the birth dates given in the birth histories. Both the numerators and the denominators of age-specific fertility rates, infant, and child mortality rates, and other important rates depend on reported age. In addition, the quality of the reports of ages and dates reflects on the quality of other information in the surveys [5]. Estimates of levels and trends in such rates may be affected by misreporting of ages and dates of birth for a woman and her children, or dates of death for her children. Age displacement of mothers and their children can seriously distort estimates of current levels and recent trends in fertility and mortality. Hence, the quality of data of these surveys, its assessment, and adjustment are of utmost importance to data users and researchers as to produce reliable estimates that can build on evidence-based policies and programs.

The completeness and quality of data on birth date and, consequently the age of child are important, because they are usually used in the calculation and interpretation of child health indicators. As earlier mentioned, the age of child is also a criterion of inclusion in the health section of the questionnaire. Some of children's birth dates were pushed back by the interviewer a year or two, so that the health section could be skipped for those children [1]. Moreover, there is a problem of misplacement of date of birth (month and year of birth) which leads to age over or under reporting, age distributions of children will often suggest a preference for even numbers [5]. Accuracy and completeness of child age are also crucial for the analysis of anthropometric data. Weight and height for age of each child is compared to weight and height of children of a reference population in the same child's age in months, only weight for height is independent of child's age. Hence, two anthropometric indicators are affected by the misreporting of the child's age, due to digital preference or an intentional push back of the date of birth by the interviewer.

2. Objectives

The objectives of this paper are to assess:

1- The quality of women's age data.

2- The children year of birth information and age in single months.

3- The quality of anthropometric data of under-five children

3. Data and Method

3.1 Data

The main source of data for this study is the Sudan Household Health Survey, 2010 (SHHS, 2010) raw data file. The SHHS, 2010 was carried out during the period March to May 2010 by the Sudan Federal Ministry of Health and the Central Bureau of Statistics (CBS). The methodology and contents of SHHS 2010 was based on the Multiple Indicators Cluster Survey (MICS), an international household survey programme developed and supported by UNICEF, and Pan Arab Project for Family Health (PAPFAM). The 15 state of Sudan constituted the main sampling domains and in each state a two stage cluster sampling design was employed to draw the sample for the SHHS, 2010 [6]. The sample was stratified by states, urban and rural. Five types of questionnaires were used in this survey, namely: Household, Female, Under-five children, Community and Food Security questionnaires.

3.2 Sample Coverage

Of the 15,000 households selected for the sample, 14,921 were found to be occupied. Of these, 14,778 households were interviewed successfully with a household response rate of 99%. In those households interviewed, 18,614 women (aged 15-49 years) were identified. Of these, 17,174 women were interviewed, yielding a response rate of 92.3% within interviewed households. In addition, 13,587 children under age five were listed in the household questionnaire. Questionnaires were completed for 13,282 of these children, corresponding to a response rate of 97.8 %. In the interviewed households, 16,448 men (age 15-49 years) were identified. Of these, only 5,573 men could be successfully interviewed, yielding a response rate of 33.9 % within interviewed households. One of the reasons for the low response rate for men was that, at the time of visit to the households by the interviewer, men in a large proportion of the households were out on work [6]

3.3 Method

The Myer's Index is used to calculate heaping in women's age. The standard summary measure of digit preference is produced by Myers' blended method which was developed by Shryock & Siegel, 1976[3]. The blending procedure adjusts for the natural tendency of a final '0' to occur more often than a final '1', and a final '1' more often than a final '2', and so on, simply because population growth and mortality tend to result in more people age x than age x+1. The blending procedure for the index requires that the age range be a multiple of ten years, so that each final digit (0 through 9) to occur same number of times [5].

The birth year ratio of children is calculated to examine the extent of displacement of children. The birth year ratio is defined as the numbers of births in year **X** relative to average number of births in year **X+1** and Year **X-1** (year succeeding and year proceeding the year of interest) multiplied by 100. The ratio is expected to be 100 in the absence of displacement, and greater than or less than 100 if there is displacement. Percent distribution of number of children by age in single month is used to detect preference of age reporting in certain months. Moreover, distributions of children weight and height measures are calculated to examine the level of digits preference in these measures.

4. Results

4.1 Heaping in Women's Age

As indicated in Table 1, Myer's Index for women ages (20-39) shows that 20.6% of women ages ended in '5', (19.6%) ended in '0'. The highest avoidance is observed for digit '1' (-5.2%) followed by digit '4' (-3.9%). The overall heaping index is 20%. Figure 1 shows the percent distribution of each of the final digits (0-9) for women ages 20-39 years to their total distribution; it reflects heaping at final digit '0' and '5' (22% and 21% respectively). This heaping in women's age can be smoothened by grouping women ages.

4.1.2 Heaping in Women's Age by Mode of Living

Myer's Index for age of women by mode of living as shown in tables 2 and figure 2 indicates disparity in age heaping between age of rural women and urban ones (24% and 15% respectively) which demonstrates the fact that urban women are more educated than their rural counterparts and are expected to have birth certification and know their ages better than rural women.

4.1.3 Heaping in Women's Age by their Educational Level

Women educational level has pronounced effect on the level of age heaping as shown in the successive tables (tables 3, 4, 5, and 6 and figure 2). As women's educational level increases, the overall heaping index decreases (10%, 18%, 29% and 30% respectively for ages of women with secondary+, primary, no education and khalwa/adult education). This emphasizes the importance role of women's education in providing accurate information for surveys and censuses data.

4.1.4 Children's Years of Birth Information

Figure 3 shows that, there is deficit in the number of births in the second and fourth years preceding the SHHS, 2010, the year of birth ratios for these years are 95.3%, and 90.4% respectively, while there is excess of births in the third year (109%), and that maybe due to the fact that interviewers intended to push births date forward to skip the question of breastfeeding which was asked for children aged less than 2 years at the date of the survey. Very slight overestimation is observed in the fifth and sixth years preceding the survey, with year of birth ratios of 103% and 102% respectively (Figure 3). The year birth ratio results indicate that there is improvement in data accuracy compared to previous surveys, for instance, comparing the birth year ratios of Sudan Maternal and Child Health Survey, 1993 and Sudan Household Health Survey 2006 for the fifth year preceding the survey (70.6% and 96.7% respectively) with 103% for SHHS, 2010, there was deficit in numbers of births in the fifth year preceding the first two surveys. Thus, that improvement in data of SHHS, 2010 may be attributed to the general awareness of people on the importance of birth registration, the rules and regulations put in place for linking birth certificate with other civil services and the establishment of the National Civil Register as well as better training for interviewers.

4.1.5 Heaping in Age of Child in Months

Household health and demographic survey's findings suggest that there is tendency to heap towards even numbers in child's age reporting and more specifically in multiples of six. Figure 4 confirms that assumption, as there is noticeable peak in the number of children at ages 12, 24, 36…..48 compared to those whose ages are in odd numbers. However, grouping of child's age is suggested to reduce this digit heaping and smoothen the data pertains to child's age. As it is apparent from figure 5, the percent distribution of age groups is almost normal with the exception

of age group 0-11 months, which is an expected pattern, as fewer births were delivered in one year than 2 years and more before the date of the survey.

4.1.6 Heaping in Height and Weight Measures

As in other measures (women, and children ages), there might be misreporting in weight and height measures, with interviewer's preference of recording certain terminal decimal digits, for instance ".0" and/or ".5". It is assumed that there should be an even distribution of the readings on each of the first ten decimal digits (0-9), that is to say 10% of the measurements should fall on each digit [4]. The evaluation of the DHS countries data explored that heaping in weight and height measures is most common at '.0' and '.5' (whole and half numbers) [4]. It is worth noting that heaping in anthropometric measures was solely an interviewer mistake, as measures were taken and reported by the interviewer and the respondent played no role in this activity. Figure 6 shows the percent distribution of final decimal digits of height readings in SHHS, 2010 data. It is apparent that heaping at '.0' (whole number) is high (32%) of the readings recorded with whole numbers, followed by the ones ending at '.5' (half numbers) (12%). Figure 7 shows the percent distribution of weight readings of terminal first decimal digits (0-9), the graph confirms that with exception of '.0' which has 2% higher than theoretical percentage (10%) and '.5' has 1% higher than the 10%, the digits distribution is almost normal, which indicates that weight measurements is more accurate than the height ones, and confirms that there is no heaping at overall level.

5. Conclusion

This paper attempts to assess some aspects of data of Sudan Household Health Survey, 2010, (SHHS, 2010) namely: information of age of women, children age and anthropometric measures.

The response rate of SHHS, 2010 is high (above 90%), in different types of questionnaires, with the exception of men questionnaire (34%), and as earlier explained , the time of the interviewer was where large proportion of men were at work.

The heaping rate in reporting of women age is moderate in SHHS, 2010. There is evidence of differentials in age heaping by women's educational level which emphasizes the importance role of women's education in providing accurate information for surveys data. Women education plays a positive role in reporting their ages.

The year birth ratio findings indicate that information on children's birth date is slightly misplaced. Findings of year birth ratios also indicate that interviewers intended to skip the

module of breastfeeding which was asked for children aged less than two years, instead of the whole under-five questionnaire, as was the case in previous household health surveys.

There is preference towards reporting age of children in even months, particularly towards the ones of multiple of six. However, grouping of child's age reduces this digit heaping and smoothen the data pertains to child's age.

It is apparent that height heaping at '.0' (whole number) is the highest of the readings, followed by the ones ending at '.5' (half numbers). Weight measurements are more accurate than the height ones, and there is no heaping at overall level.

6. Recommendations

The interview time may induce some misreporting, as evidence in men interview; also late time may distort the quality of interview. Hence, good interview timing is of high importance to surveys data quality.

Advancing new technology in interviewers training saves interview time and minimizes interviewer’s errors.

Appraising misreporting and heaping is critical for fertility, infant, child mortality, anthropometric measures and other child health indicators.

Women education plays an important role in accuracy of reporting their ages.

7. References

[1] Bicego George & Boerma J, Demographic Aspects of Quality of Date of Birth in the Demographic and Health Surveys, An Assessment of the Quality of Health Data in DHS-1Surveys. Methodological Report 2, Calverton, Maryland: Macro International Inc, USA, 1993.

[2] El Jack Hagir, Interviewer effects: Evidence from Sudan Health Households surveys

(1992-2010). World Association for Public Opinion Research 2015 Regional Conference on "Innovation in Public Opinion Research". Doha, Qatar7-9 March, 2015.

[3] Shryock H, & Jacob Siegel, The Methods & Materials of Demography. Academic Press, New York, USA, 1976.

[4] Sommerfelt, A. Elizabeth, and S. Kathryn, Children's Nutritional Status. DHS Comparative Studies No. 12. Macro International Inc., Calverton, Maryland, USA, 1994

[5] T.W. Pullum, A Statistical Reformulation of Demographic Methods to Assess the Quality of Age and Date Reporting, with Application to the Demographic and Health Surveys, 2005.

[6] The Sudan Household Health Surve, 2010 Report. Sudan Federal Ministry of Health, and Central Bureau of Statistics. Khartoum, Sudan, 2012.

Table 2 Myer's Index for Women's Age in Urban Areas

7.Tables

Table 1 Myers' blended Index of Women's Age Heaping, SHHS 2010

Terminal Digit, a	Number of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of Percentage from 10%	Remarks
	20+a	30+a	coefficient	coefficient						
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8		
0	1148	1105	1	9	1148	9945	11093	19.6	9.6	Preference
1	431	232	2	8	862	1856	2718	4.8	-5.2	Avoidance
2	705	449	3	7	2115	3143	5258	9.3	-0.7	Avoidance
3	523	251	4	6	2092	1506	3598	6.4	-3.6	Avoidance
4	479	213	5	5	2395	1065	3460	6.1	-3.9	Avoidance
5	1218	1096	6	4	7308	4384	11692	20.6	10.6	Preference
6	463	271	7	3	3241	813	4054	7.2	-2.8	Avoidance
7	578	369	8	2	4624	738	5362	9.5	-0.5	Avoidance
8	592	397	9	1	5328	397	5725	10.1	0.1	Ideal
9	368	325	10	0	3680	0	3680	6.5	-3.5	Avoidance
Total	6505	4708					56640	100.0	40.5	

Overall Myer's Index = 40.5%/2= 20.3%

Source: Computed from the Raw Data File of the SHHS, 2010

Terminal Digit, a	No of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of Percentage from 10%	Remarks
	20+a	30+a	coefficient	coefficient						
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8		
0	332	325	1	9	332	2925	3257	17.0	7.0	Preference
1	158	86	2	8	316	688	1004	5.3	-4.7	Avoidance
2	229	175	3	7	687	1225	1912	10.0	0.0	Ideal
3	208	99	4	6	832	594	1426	7.5	-2.5	Avoidance
4	178	93	5	5	890	465	1355	7.1	-2.9	Avoidance
5	302	323	6	4	1812	1292	3104	16.2	6.2	Preference
6	163	97	7	3	1141	291	1432	7.5	-2.5	Avoidance
7	220	119	8	2	1760	238	1998	10.5	0.5	Preference
8	222	139	9	1	1998	139	2137	11.2	1.2	Preference
9	149	129	10	0	1490	0	1490	7.8	-2.2	Avoidance
Total							19115	100.0	29.7	

Overall Myer's Index for age of urban women = 29.7%/2 = 15%

Source: Computed from the Raw Data File of the SHHS, 2010

Table 3 Myer's Index for Women's Age in Rural Areas

Terminal Digit, a	No of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of Percentage from 10%	Remarks
	20+a	30+a	coefficient	coefficient						
a	**1**	**2**	**3**	**4**	**5 =(1*3)**	**6=(2*4)**	**7=(5+6)**	**8**		
0	816	780	1	9	816	7020	7836	20.9	10.9	Preference
1	273	146	2	8	546	1168	1714	4.6	-5.4	Avoidance
2	476	274	3	7	1428	1918	3346	8.9	-1.1	Avoidance
3	315	152	4	6	1260	912	2172	5.8	-4.2	Avoidance
4	301	120	5	5	1505	600	2105	5.6	-4.4	Avoidance
5	916	773	6	4	5496	3092	8588	22.9	12.9	Preference
6	300	174	7	3	2100	522	2622	7.0	-3.0	Avoidance
7	358	250	8	2	2864	500	3364	9.0	-1.0	Avoidance
8	370	258	9	1	3330	258	3588	9.6	-0.4	Avoidance
9	219	196	10	0	2190	0	2190	5.8	-4.2	Avoidance
Total							37525	100.0	47.5	

Overall Myer's Index for Age of Rural women=47.5%/2=23.8%

Source: Computed from the Raw Data File of the SHHS, 2010

Table 4 Myer's Index of Age of women with No Education

Terminal Digit, a	Number of women ages 20-39		Weight for	Blended Population		Percent Distribution	Deviation of Percentage from 10%		Remarks
	20+a	30+a	coefficient	coefficient					
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8	
0	421	504	1	9	421	4536	4957	22.8	12.8	Preference
1	95	76	2	8	190	608	798	3.7	-6.3	Avoidance
2	213	175	3	7	639	1225	1864	8.6	-1.4	Avoidance
3	120	112	4	6	480	672	1152	5.3	-4.7	Avoidance
4	118	67	5	5	590	335	925	4.3	-5.7	Avoidance
5	563	559	6	4	3378	2236	5614	25.9	15.9	Preference
6	153	102	7	3	1071	306	1377	6.3	-3.7	Avoidance
7	200	167	8	2	1600	334	1934	8.9	-1.1	Avoidance
8	206	175	9	1	1854	175	2029	9.4	-0.6	Avoidance
9	105	124	10	0	1050	0	1050	4.8	-5.2	Avoidance
Total							21700	100.0	57.4	

Overall Myer's Index for age of women with no education=28.7%

Source: Computed from the Raw Data File of the SHHS, 2010

Table 5. Myer's Index for Age of Women with Khalwa/Adult Education

Terminal Digit, a	Number of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of Percentage from 10%	Remarks
	20+a	30+a	coefficient	coefficient						
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8		
0	57	75	1	9	57	675	732	27.0	17.0	Preference
1	14	6	2	8	28	48	76	2.8	-7.2	Avoidance
2	36	15	3	7	108	105	213	7.9	-2.1	Avoidance
3	18	10	4	6	72	60	132	4.9	-5.1	Avoidance
4	14	13	5	5	70	65	135	5.0	-5.0	Avoidance
5	65	60	6	4	390	240	630	23.3	13.3	Preference
6	14	21	7	3	98	63	161	5.9	-4.1	Avoidance
7	28	14	8	2	224	28	252	9.3	-0.7	Avoidance
8	25	22	9	1	225	22	247	9.1	-0.9	Avoidance
9	13	18	10	0	130	0	130	4.8	-5.2	Avoidance
Total							2708	100.0	60.6	

Overall Myer's index for age of women with khalwa/adult education=30%

Source: Computed from the Raw Data File of the SHHS, 2010

Table 6 Myer's Index for Age of Women with Primary Education

Terminal Digit, a	Number of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of Percentage from 10%	Remarks
	20+a	30+a	coefficient	coefficient						
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8		
	386	311	1	9	386	2799	3185	18.0	8.0	Preference
1	132	72	2	8	264	576	840	4.7	-5.3	Avoidance
2	232	152	3	7	696	1064	1760	9.9	-0.1	Avoidance
3	167	65	4	6	668	390	1058	6.0	-4.0	Avoidance
4	160	64	5	5	800	320	1120	6.3	-3.7	Avoidance
5	360	320	6	4	2160	1280	3440	19.4	9.4	Preference
6	157	94	7	3	1099	282	1381	7.8	-2.2	Avoidance
7	187	112	8	2	1496	224	1720	9.7	-0.3	Avoidance
8	199	122	9	1	1791	122	1913	10.8	0.8	Preference
9	129	108	10	0	1290	0	1290	7.3	-2.7	Avoidance
Total							17707	100.0	36.5	

Overall Myer's Index for age of women with primary education = 18.3%

Source: Computed from the Raw Data File of the SHHS, 2010

Table 7 Myer's Index for Age of Women with Secondary and above education

Terminal Digit, a	Number of women ages 20-39		Weight for		Blended Population			Percent Distribution	Deviation of percentage from 10%	
	20+a	30+a	coefficient	coefficient						
a	1	2	3	4	5 =(1*3)	6=(2*4)	7=(5+6)	8		
0	284	215	1	9	284	1935	2219	15.3	5.3	Preference
1	190	78	2	8	380	624	1004	6.9	-3.1	Avoidance
2	224	107	3	7	672	749	1421	9.8	-0.2	Avoidance
3	218	64	4	6	872	384	1256	8.6	-1.4	Avoidance
4	187	69	5	5	935	345	1280	8.8	-1.2	Avoidance
5	230	157	6	4	1380	628	2008	13.8	3.8	Preference
6	139	54	7	3	973	162	1135	7.8	-2.2	Avoidance
7	163	76	8	2	1304	152	1456	10.0	0.0	Ideal
8	162	78	9	1	1458	78	1536	10.6	0.6	Preference
9	121	75	10	0	1210	0	1210	8.3	-1.7	Avoidance
Total							14525	100.0	19.5	

Overall Myer's Index for age of women with secondary education = 10%

Source: Computed from the Raw Data File of the SHHS, 2010

8. Figures Caption

Figure 1 Percent Distribution of Final Digits of Women Aged 20-39, SHHS, 2010

Figure 2 Myer's Blended Index for Women Aged 20-39, SHHS, 2010 by Mode of Lining and Educational Level

Figure 3 Number of Births by Calendar Years Preceding the Survey, SHHS 2010.

Figure 4 Distribution of Number of Children for Child's Age in Single Month for Children aged 0-50 Months, SHHS, 2010

Figure 5 Percent Distribution of Children Age Groups for Children Aged 0-59 months, SHHS, 2010.

Figure 6 Percent Distribution of Terminal Decimal Digits for Height Readings, SHHS, 2010

Figure 7 Percent Distribution of Terminal Decimal Digits for Weight Readings, SHHS, 2010

9. Figures

Figure 1. Percent Distribution of Final Digits for Women aged 20-39 years, SHHS,2010

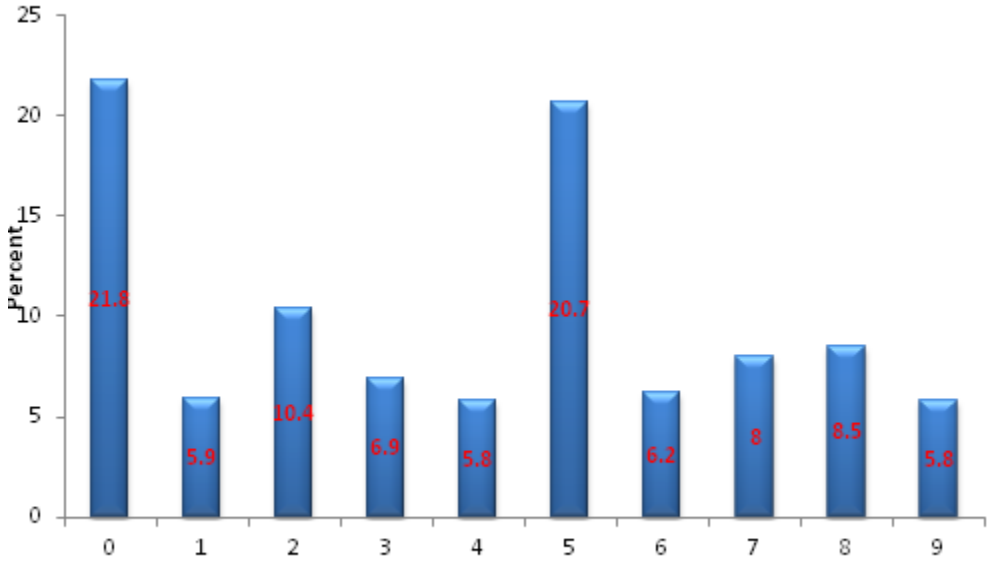
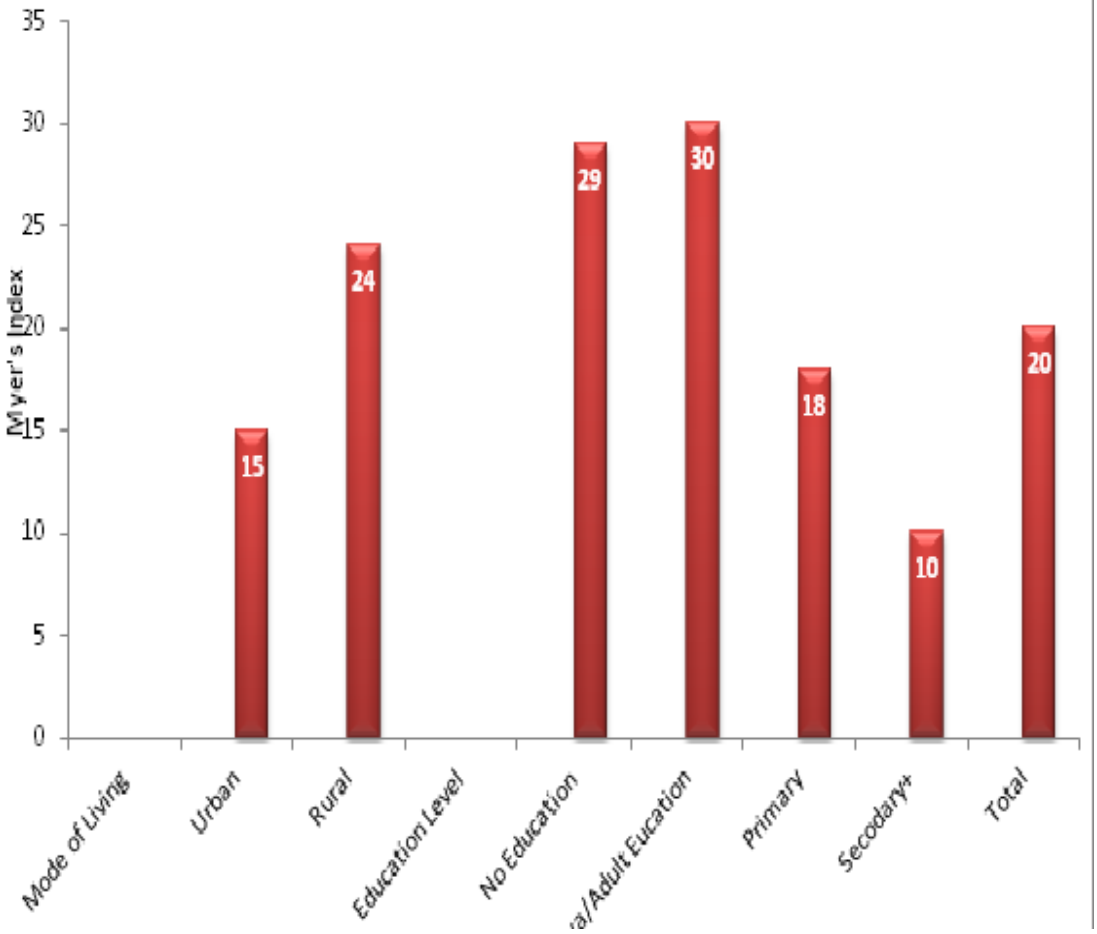


Figure 2. Myer's Blended Index for Women Ages 20-39 years by Mode of Living and Education Level, SHHS, 2010



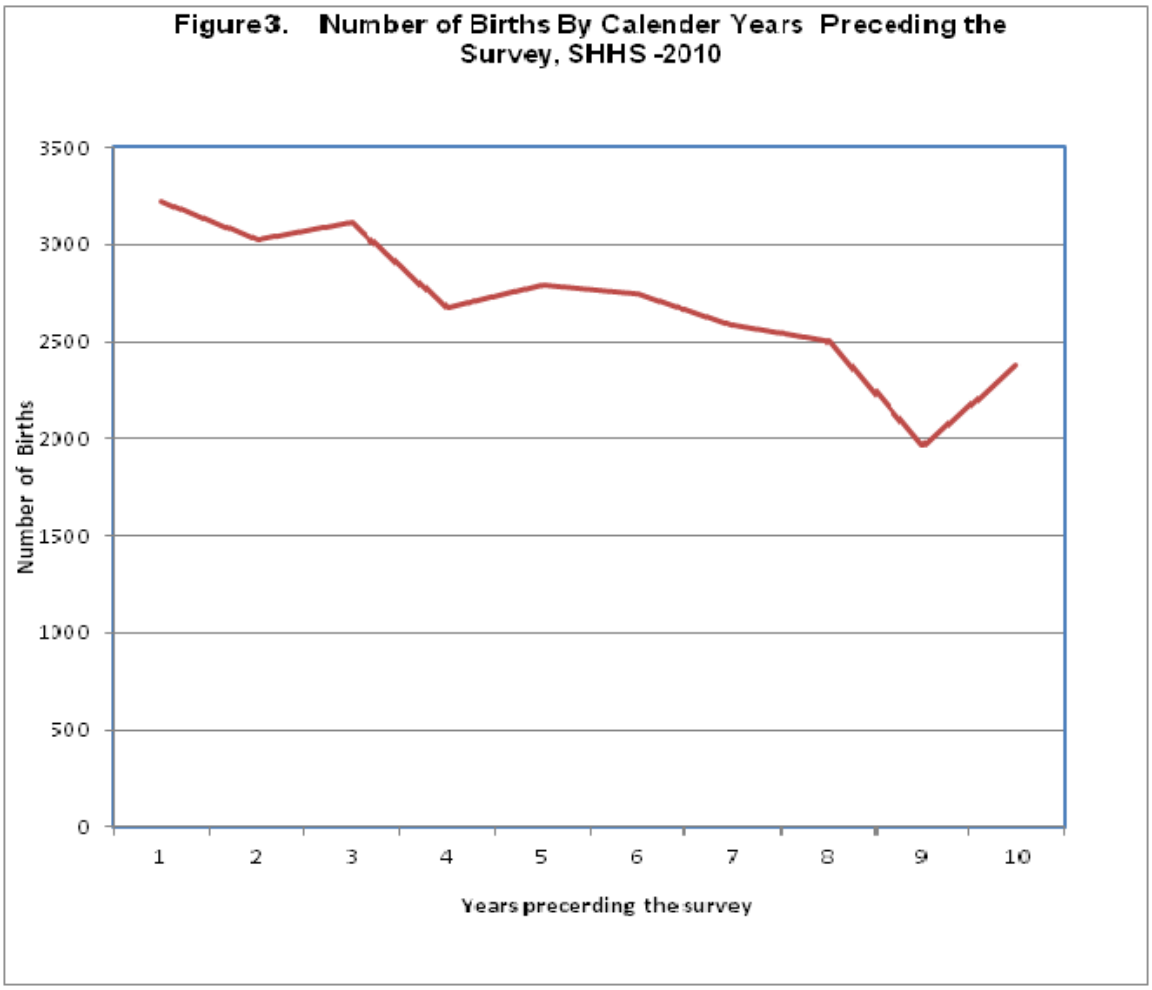


Figure3. Number of Births By Calender Years Preceding the Survey, SHHS -2010
3500
3000
2500
2000
1500
1000
500
0
Number of Births
1
2
3
4
5
6
7
8
9
10
Years precerding the survey

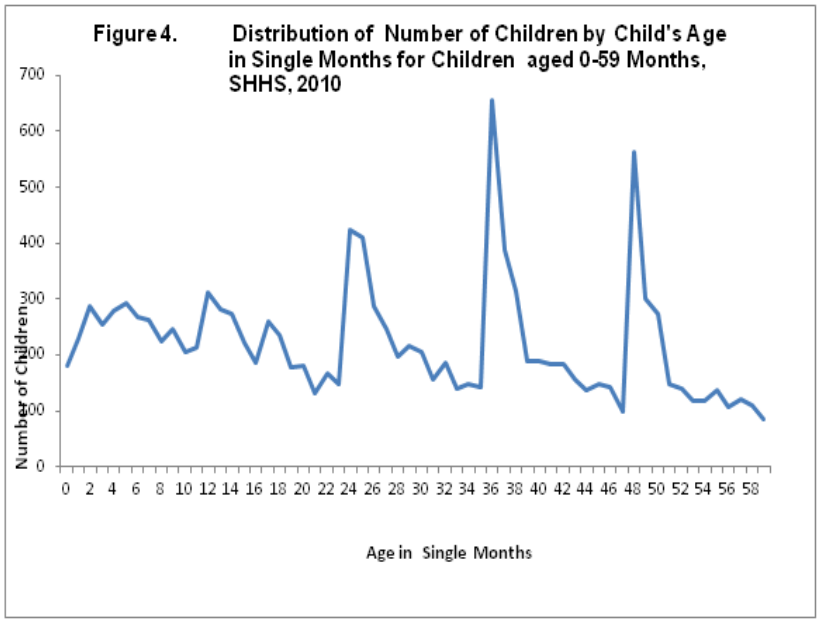
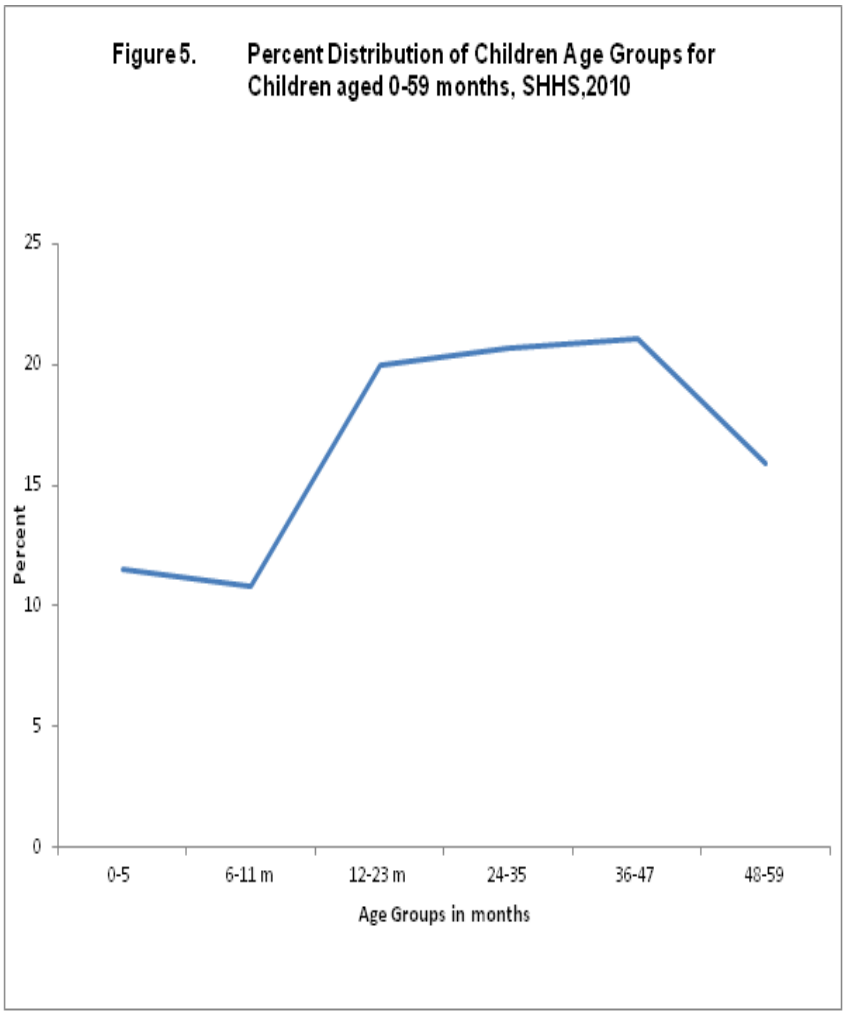


Figure 4. Distribution of Number of Children by Child's Age in Single Months for Children aged 0-59 Months, SHHS, 2010
Number of Children
700
600
500
400
300
200
100
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58
Age in Single Months

Figure 5. Percent Distribution of Children Age Groups for Children aged 0-59 months, SHHS,2010



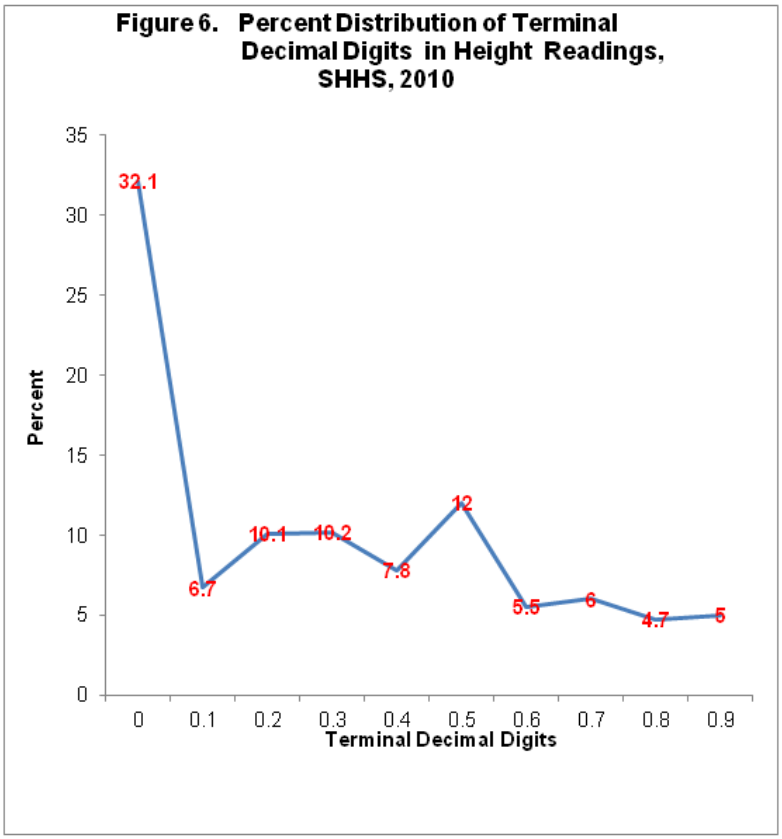


Figure 6. Percent Distribution of Terminal Decimal Digits in Height Readings, SHHS, 2010
Percent
35
30
25
20
15
10
5
0
32.1
6.7
10.1
10.2
7.8
12
5.5
6
4.7
5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Terminal Decimal Digits

Figure 7. Percent Distribution of Final Decimal Digits in Weight Readings, SHHS, 2010