

Data quality-Comparison between Census and DSS outputs for Navrongo-Ghana

Introduction

Most countries of the developing world lack accurate and complete demographic data for any meaningful analysis (Mahapatra et al. 2007; 2005; Soto et al, 2013). This paucity of data has been due to the lack of well-established vital registration systems in these countries (United Nations, 1998). Traditional sources of demographic data in Africa are population censuses and health and demographic surveys, and these sources have drawbacks, which include the problem of recall bias which sometimes results in the wrong estimation of some rates (Byass et al, 2007).

An important system of demographic data collection in developing countries that is now gaining popularity is demographic surveillance systems (DSSs). A number of these DSSs sites have been established across Africa and Asia. One of such DSS is Navrongo Demographic Surveillance System (NDSS), which is located in two districts of northern Ghana.

Ghana conducted a census in 2010 with the reference date being 26th September 2010. The census report characterized the socio-demographic profile of each district of the country including the Kassena-Nankana Municipal and Kassena-Nankana West districts where the Navrongo DSS operates.

The two data sources (census and NDSS) offer the opportunity to validate the accuracy and reliability of each data collection system. The objective of this paper is therefore to examine the common output from the two data sources to see how they compare with each other.

Data source

Data for this analysis come from two independent sources, the 2010 Population and Housing Census (PHC) and the NDSS. In 2010 Ghana conducted a population and housing census with a reference date of 26th September. The Ghana Statistical Service is the institution in charge of the conduct of the census. Details of the census data collection processes can be found elsewhere (GSS, 2013). Under the HDSS system, fieldworkers visit all households every four months to collect and update the health and socio-demographic information of the people and

their households. Demographic data collected include pregnancies, births, deaths, migrations, among others (Oduro et al, 2012).

Method of analysis

The main areas of comparison are population size, age-sex distribution and mortality levels of the study area. Population pyramids, line and bar graphs are used for the comparison. T-test is used to check for any significant difference between the outputs from the two systems.

Results

We first compare the population characteristics as enumerated by the census and the DSS. The population of the study area (Navrongo) as recorded by the census was 180,611, while that of the HDSS recorded a population of 155,112. The sex ratio is 96.1 and 91.2 as recorded by the census and the DSS respectively. While the census had 22.1 percent of the population residing in the urban area the DSS had 22.2 percent urban.

Figure 1 below shows a graph of the age distribution of the population reported in the census and the DSS, with the census having proportionately more children than the DSS up to age 9 years. The population distribution for the two systems matches up to age 49 years and then the DSS proportions become slightly higher than the census proportion.

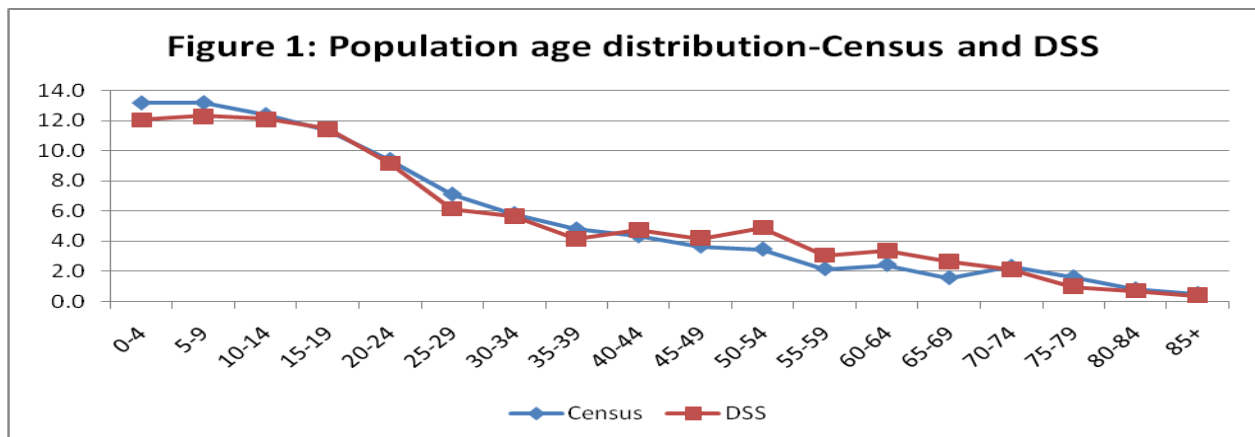
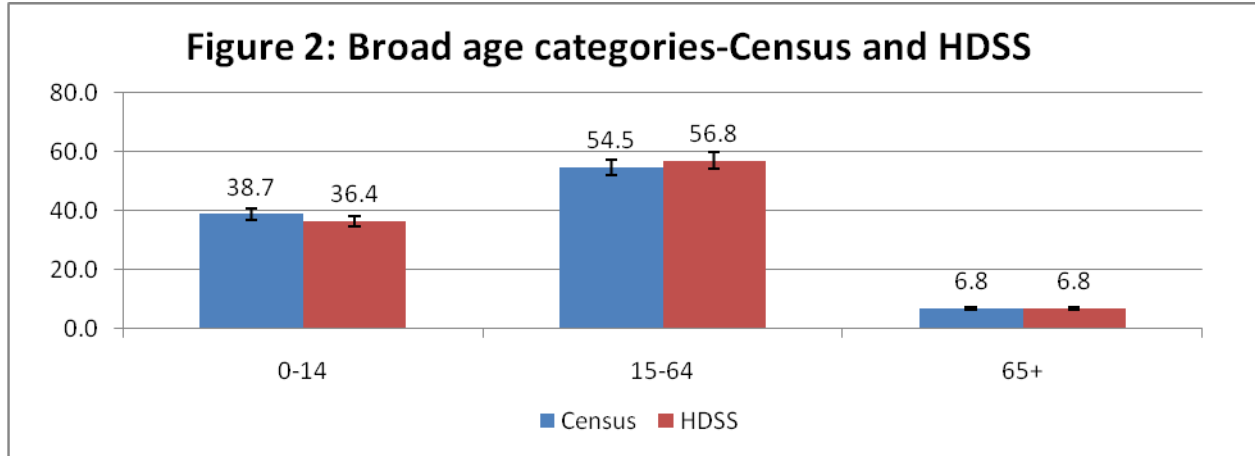
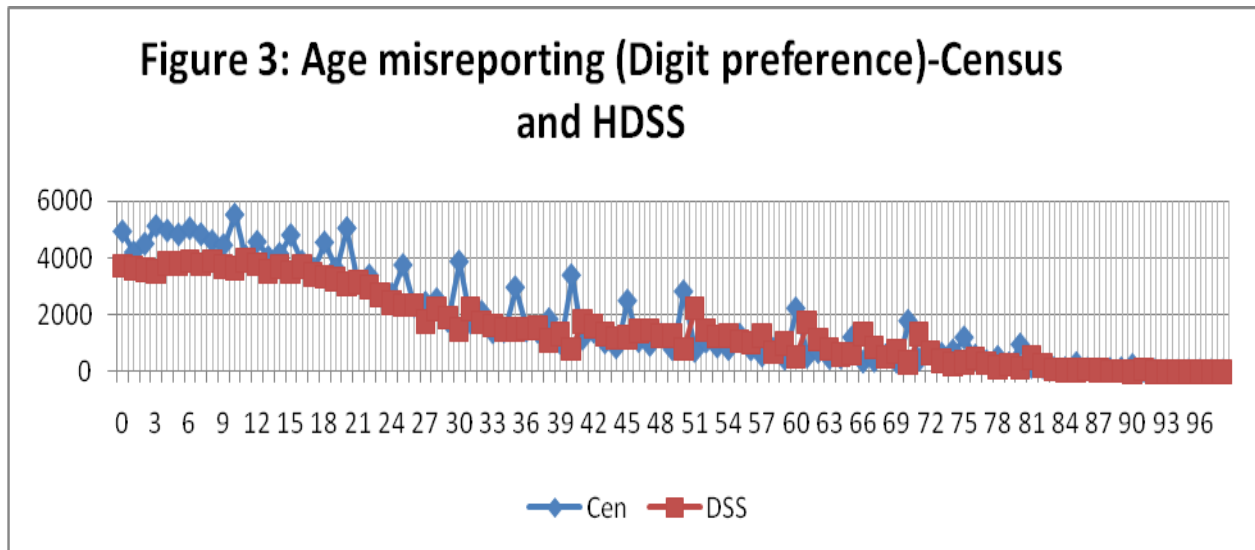


Figure 2 shows a graph of the broad age categorization of the population as recorded by the census and the HDSS. The graph shows no significant difference in the percentage distribution of the broad age categorization between the census and the HDSS population.

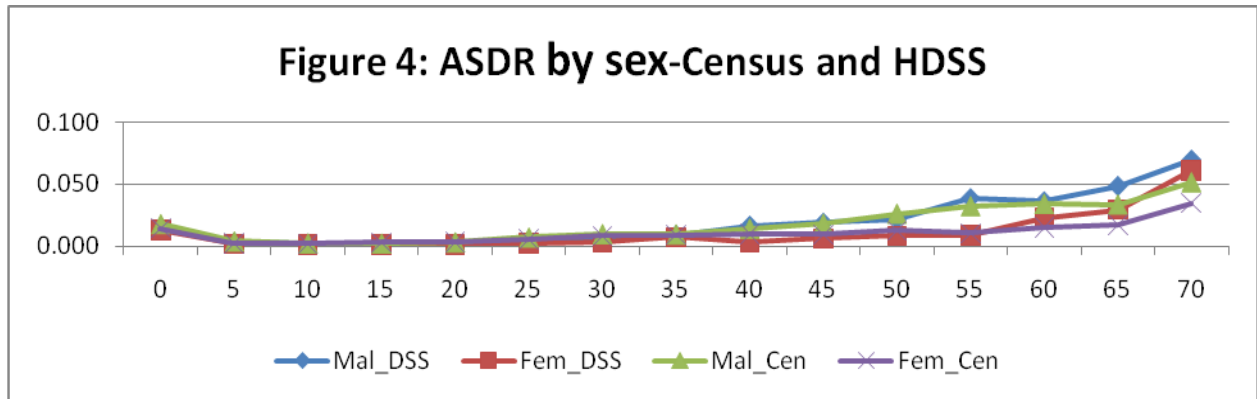


Age is an important variable in any analysis. However, despite its importance, it is one of the difficult variables to collect and analyze in terms of its accuracy and reliability. Figure 3 shows a graph of digit preference for DSS and Census. It is clear that the two data collection systems have problems with age misreporting (digit preference) but the DSS tends to have better age reporting than the census.



Mortality

Total deaths recorded within one year up to the reference date were 1783 and 1546 giving crude death rates of 9.9 and 10.0 per 1000 for the census and HDSS respectively. Figure 6 shows the age-specific death rates by sex for Census and HDSS. The expected pattern is observed for the two systems, with some little variations, particularly at age 40 and beyond.



Conclusion

The results have shown some consistency between the census and DSS. Doing a comparative analysis of this type to validate two demographic data collection systems gives confidence or otherwise to data users and shows the level of quality and reliability of the data. On the other hand if there are marked variations in the results, measures can be put in place to identify the source of these discrepancies and find ways of addressing them in subsequent waves of data collection.