

Reliability of National Data Sets: Evidence from a Detailed Small Area Study in Rural Kathmandu Valley, Nepal

¹⁻³P. Simkhada, ^{2,3}E. van Teijlingen, ⁴S. Kadel,

⁵J. Stephens, ¹S. Sharma and ⁴M. Sharma

¹Institute of Applied Health Science, University of Aberdeen, Aberdeen, AB25 2ZD, Scotland, UK

²School of Health and Social Care, Bournemouth University, England, UK

³Manmohan Memorial Institute of Health Sciences, Purbanchal University, Kathmandu, Nepal

⁴Green Tara Nepal, Kathmandu, Nepal

⁵Green Tara Trust UK, London, UK

Abstract: Researchers often rely on census data to provide us with information for local areas. In a study, we came across major discrepancies in rural Nepal between the number of women with a child under the age of two as estimated from the national census and the prevalence rate of this population in our local in-depth household survey. This study highlights why census data might not be as reliable as one would hope. In summary, researchers using census data in developing countries should include an element of quality control of the national dataset. We advise researchers to conduct a small survey from a random sample to provide an estimate of the likely population in the area under study.

Key words: Census, reliability, population estimate, measurement, data collection, household survey

INTRODUCTION

There have been calls to improve global health statistics (Boerma and Stansfield, 2007) including the need to monitor progress on health targets such as those in the millennium development goals. The inability to generate reliable information needed to make decisions is a major obstacle to healthcare planning in many developing countries (WHO, 2005). Health policy makers, planners and researchers in many developing countries without large volumes of routinely collected national demographic, health and health services data have to rely on national census data and local estimates derived such nation-wide surveys. Usually, a national census makes a huge variety of general statistical information on society available to policy-makers and researchers, but because of its size (nation-wide) it is expensive and therefore is often held with large time intervals in between. In Nepal the national census is held every ten years, collecting data door-to-door from every household in the country and the most recent one was held in 2001 (Central Bureau of Statistics, 2001). This means that the most up-to-date information can be over a decade old, depending on when the Census analysis becomes available to researchers.

Corresponding Author: Dr. Padam Simkhada, Institute of Applied Health Science, University of Aberdeen, Aberdeen AB25 2ZD, Scotland, UK
Tel: 0044-1224-552495

In developed countries, routinely collected patient demographics are available. In Nepal, as in many developing countries, the national DHS (Demographic Health Survey) type health data (Ministry of Health Nepal, New ERA, ORC Macro, 2002; DHS, 2007) together with other surveys of a sample of a proportion of the total population provide the necessary data that informs policy. Although the quality of self-reported health data, especially the data collected amongst the poorest part of the population has been questioned in the literature (Sen, 2002; King *et al.*, 2004), it is often the best data available. Data validity and reliability is also discussed when analysing DHS household survey in which women of reproductive age are interviewed regarding their recent births including Caesarean Section, it is argued that as the DHS surveys tend to produce higher caesarean section rates than data from the health facilities that the data is insufficiently precise (Stanton *et al.*, 2005). The validity and reliability of census data has rarely been examined. This study highlights why census data might not be as reliable as one would hope.

MATERIALS AND METHODS

As part of a long-term maternity-care study in Nepal, we collected baseline health and demographic data in four Village Development Committees (VDCs) in Kathmandu District in January and February 2008. We have anonymised the name of VDCs to A, B, C and D. These four are typical VDCs in Kathmandu Valley which are relatively underdeveloped, but slightly more developed than the average VDC in more remote parts of rural Nepal.

Of the four VDCs in our study, A and B are situated 20 to 25 km East of Kathmandu. The 2001 Census suggested that there were 4,417 people in VDC A and 3,880 in VDC B. Village D is 20 km South of Kathmandu. According to the national Census of 2001 there are a total of 824 households while the total population of a VDC was 4,427; about half of them female (Central Bureau of Statistics, 2001). Some of the wards in VDC D are connected by road to Kathmandu. VDC C is 3 km from VDC D and the number of households was similar to that in VDC D. VDC C has a total population of 4,142 and slightly more than half were women (Central Bureau of Statistics, 2001). All four VDCs contained nine wards each. As this study focused on maternity care we were interested in studying women with at least one child under the age of two. The census informs us that a national level 4.2% of the total population constitutes women with at least one child under the age of two.

Based on the national census, we estimated the total sample (population) for the four VDCs to be 708 women with a child under the age of two ($4.2\% * 16,866$). In 2008, using trained Nepalese fieldworkers we subsequently visited every household in each VDCs over a two-month period to collect baseline information.

RESULTS

Having visited all households in all four VDCs we could only find evidence of 485 women with a child under the age of two. Of these 485 women 412 women agreed to complete our survey, 36 declined to participate and 37 could not be found despite several visits to their homes. The women who refused to participate or who could not be found were reasonably well distributed across the four VDCs. Socio-demographic information of women who agreed to take part in survey is shown in Table 1. The single largest group consisted of Tamang women (39%). The majority were younger than 25 years old, most fitted the description of housewife and over one third was illiterate.

Table 1: Socio-demographic information of women (N = 412)

Characteristics	Frequency (n)	Percentage
Caste of women		
Brahmin/Chhetri	134	33
Tamang	161	39
Newar	88	21
Balami/Dalit/Others	29	7
Age of women		
15-24	218	53
25-34	173	42
35-44	21	5
Occupation of women		
Housewife	287	70
Farmer	98	24
Service/Business	27	6
Literate		
Yes	269	65
No	143	35

DISCUSSION

According to the 2001 census one would have expected to find 708 mothers in the present study area, however, from our 2008 comprehensive household survey we could only find 485. In other words, 223 women (708 minus 485) were missing, which means 31% of women with a child under the age of two had disappeared, in the span of 7 years. There are several possible logical explanations for this discrepancy, namely (1) both data sets are right but there has been a change in the population over time; (2) the way Census data was amalgamated introduced anomalies; (3) the Census data are imprecise or incorrect and (4) our data are incorrect. Finally, an issue we will address here, there is, of course the possibility that both data sets are incorrect.

Change Over Time

It is possible that the difference between the census and our study is a reflection of reality. Perhaps the missing women in this category have emigrated or the death rate has been higher than the national average between the time the 2001 census was held and our 2008 survey. This is, of course a possibility as there are more than 3 million Nepalese working abroad (Kollmair *et al.*, 2006) and there is internal migration in Nepal towards the capital.

Anomalies in the Census Data

Is it possible that the national proportion of women in this category is different from that in our survey as the population is somehow skewed. We found that the data estimation of the VDCs is based on an amalgamation of 57 VDCs (out of about 4,000 nation-wide), one metropolitan city and one municipality in Kathmandu District and proportions are based on a straightforward division of the total population in the district into each VDC. As this district includes the large urban centre of the capital these proportions may not be applicable for the more rural areas. The official statistical predictions made for 2008 for each VDC are based on the 2001 census. However, this does not take into consideration the high levels of growth in the city of Kathmandu due to internal migration. This migration affects the proportion of women of child bearing age, partly because the population influx into the capital does not reflect the national proportions of sub-groups. For example, the literature suggests that young men are more likely to migrate than either young women or older men (Bhattarai, 2005).

Imprecise Census Data

The census in 2001 was imprecise due to the internal conflict and subsequent poor data collection. Due to the violent conflict at the time between Maoist rebels and the Nepalese Government (Devkota and van Teijlingen, 2007) data collection in some parts of country would have been too dangerous.

Our Data is Incorrect

Although, always a possibility, we feel that this explanation is unlikely. We conducted a local in-depth survey at a level of precision, such as revisiting homes where we knew from neighbours that a woman with a child under the age of two lived, which is hard to achieve at a national level in the census.

Thirty one percent missing women can be attributed to migration (internal and external), population growth, internal conflict and subsequent poor data collection, each of these elements which can affect the reliability of Census data.

CONCLUSIONS

Information from population censuses can provide statistics for local areas (Hakim, 1997: 53). However, official data sets, collected using a national survey approach, have limited use at the local level if these local data estimates are being based on proportional/arithmetical representations. Policy-makers, planners and researchers need to consider the limitations and the quality of the data set before basing their decisions on such data.

We feel that general changes in the population such as internal migration and emigration may have been accelerated by the internal decade-long conflict in Nepal. Moreover, this conflict will have made data collection for the Census less reliable, since (1) part of the country was not under Government control and (2) Census enumerators might have been afraid to approach people whom they believed to be Maoist sympathisers as census enumerators were working for the government.

We suggest that as a measure of quality control, one conducts a small survey with a random sample to establish an estimate of the likely population in the area under study and in the desired target group, be it children under 5, women or reproductive age or men over 50.

REFERENCES

- Bhattarai, P., 2005. Migration of Nepalese youth for foreign employment: problems and prospects: A review of existing government policies and programmes. Youth Action of Nepal, Kathmandu. www.youthaction.org.np/downloads/Research%20Report%20on%20Labor%20Migration.pdf.
- Boerma, J.T. and S.K. Stansfield, 2007. Health statistics now: Are we making the right investments?. *Lancet*, 369: 779-786.
- Central Bureau of Statistics, 2001. Statistical Yearbook of Nepal 2001. Government of Nepal, National Planning Commission Secretariat, Kathmandu, Nepal.
- Devkota, B. and E. van Teijlingen, 2007. Basic health as peace dividend in post-conflict Nepal. *J. HEPASS*, 3: 21-23.
- DHS., 2007. Nepal demographic and health survey 2006. Ministry of Health and Population Government of Nepal, New ERA Kathmandu Nepal and Macro International Inc. USA.

- Hakim, C., 1997. *Research Design: Strategies and Choices in the Design of Social Research*. Routledge, London.
- King, G., C.J.L. Murray, J.A. Salomon and A. Tandon, 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am. Pol. Sci. Rev.*, 98: 191-207.
- Kollmair, M., S. Manandhar, B. Subedi and S. Thieme, 2006. New figures for old stories: migration and remittances in Nepal. *Migr. Lett.*, 3: 151-160.
- Ministry of Health Nepal, New ERA, ORC Macro, 2002. *Nepal demographic and health survey 2001*. Family Health Division, Ministry of Health; New ERA; ORC Macro, Calverton, Maryland.
- Sen, A., 2002. Health: Perception versus observation. *Brit. Med. J.*, 324: 860-861.
- Stanton, C., D. Dubourg, V. De Brouwere, M. Pujades and C. Ronsmans, 2005. Reliability of data on caesarean sections in developing countries. *Bull. World Health Org.*, 83: 449-455.
- WHO., 2005. Getting the numbers right. *Bull. World Health Org.*, 83: 561-640.