Medicine and Society

Using the Indian National Sample Survey data in public health research

ANURADHA KATYAL, PRABAL V. SINGH, AMIT SAMARTH, SOFI BERGKVIST, MALA RAO

INTRODUCTION

The National Sample Survey (NSS), instituted in 1950, was the brainchild of Professor Mahalanobis, widely regarded as the father of Indian statistics.[1] His ambition was to obtain and quantify comprehensive information on an annual basis on the socio-economic, demographic, agricultural and other profiles of the country, both at the national and state levels. The NSS is a multi-stage, multi-subject and multi-purpose cross-sectional survey, which is conducted annually and covers topics of current interest.[2]

The NSS data are collected and disseminated by the National Sample Survey Organization (NSSO) of India, which is a part of the Ministry of Statistics and Programme Implementation (MOSPI). It is recognized that the NSS data provide important information on the sociodemographic trends for the past 62 years.[3,4] Researchers and organizations ranging from the WHO, World Bank and Planning Commission of India to academic and private institutions use these data. The data have found their way into peer-reviewed papers published in journals, as well as blogs, newspapers and press releases. Yet the NSS methodology has probably never been replicated in designing surveys by researchers. In addition to the data published in its reports, the NSSO also has unit-level (household and individual) data, which are a vast source of information for those conducting research and making policy in the sphere of public health.

The use of NSS data is not without its challenges, as the survey architecture may seem complex to researchers who lack experience in its use. Many Indian researchers have started blogs that teach their peers how to analyse and extract the NSS data. The objective of this paper is to inform the reader about the planning and execution of the NSSs and their potential use in public health research. In doing so, we draw on our experience of using these data. The surveys are conducted throughout the country, using a sampling technique that ensures representativeness.[1,2] Hence, the sampling methodology is described in detail to enable readers to understand it and replicate it in research. Further, as there appears to be limited experience in the translation of raw unit-level data into formats that can be used for research, this paper describes how such translation can be achieved to enable researchers to make greater use of this rich source of data.

ACCESS Health International Inc., 845 UN Plaza, Suite 86A, New York, New York, USA 10017-3536
ANURADHA KATYAL, SOFI BERGKVIST

University of East London, IHHD, UH250, Stratford Campus, University of East London, Water Lane, London E15 4LZ, UK
MALA RAO Department of International Health

Oxford Policy Management, J-25 Jor Bagh Lane, New Delhi 110003, India
PRABAL V. SINGH

SughaVazhvu Healthcare, A2 L.P Amsavalli Illam, Arulananda Nagar, Thanjavur 613007, Tamil Nadu
AMIT SAMARTH

Correspondence to ANURADHA KATYAL, 103, Medha Towers, Attapur, RK Nagar, Hyderabad 500048, Andhra Pradesh; *anuradha.katyal64@gmail.com*

HOW NSSs ARE PLANNED AND EXECUTED

The MOSPI is responsible for identifying the topics of contemporary interest that the surveys should focus on. Broadly, the topics may be grouped under five areas: employment–unemployment and consumer expenditure; unorganized enterprises in non-agricultural sectors; a wide range of healthcare indicators, such as population, births, deaths, disability, morbidity, fertility, maternity and child care, and family planning; land-holdings and livestock enterprises; and lastly, debt, investment and capital formation.[3] In addition, a survey of annual consumer expenditure, of a 'thin' sample of households, is conducted.[3] The topics covered by the NSSs over the past 8 years (available online) are listed in Table I.

A survey period is referred to as a round and to date, 69 rounds have been conducted. Reports have been published up to the 66th round, and the reader can easily access these from the microdata archive on the MOSPI website[3] or by making a simple web portal search.

Sampling and data collection

The planning of the surveys ensures equity of representation across the national socioeconomic, demographic and geographical landscape. For each state or union territory, the samples are

TABLE I. Topics covered by previous rounds of the National Sample Survey

Period	Round	Subjects covered
Jan 2004–June 2004	60th	Household consumer expenditure, employment and unemployment, morbidity and healthcare
July 2004–June 2005	61st	Household consumer expenditure, employment and unemployment
2005–2006	62nd	Household consumer expenditure, employment and unemployment, manufacturing enterprises
2006–2007	63rd	Consumer expenditure, service sector enterprises (excluding trade)
2007–2008	64th	Consumer expenditure, employment and unemployment and migration particulars, expenditure on education
2008–2009	65th	Domestic tourism, housing conditions, slum particulars
2009–2010	66th	Consumer expenditure, employment and unemployment
2010–2011	67th	Un-incorporated non-agricultural enterprises
2011–2012	68th	Consumer expenditure, employment and unemployment

Source: Reference 1

divided into two sectors (rural and urban) in proportion to the provisional population, as per the Census.[5]

The ministry has a working group including statisticians and economists. The working group is responsible for the planning of the survey, including the survey design, concepts and definitions to be used, design of the questionnaire, instructions to be given to the field staff, and details of data collection and analysis.[3]

FIRST-STAGE SAMPLING

Villages and urban blocks

A stratified, multi-stage sampling method is used. The first-stage units (FSUs) consist of villages in the rural sector and urban frame survey (UFS) blocks in the urban sector. The villages in which the survey is to be undertaken are selected by a probability proportional to size with replacement method, from among the villages listed in the Census.[3]

The urban FSUs or UFS blocks are 'mapped' by the NSSO, taking into consideration the increase or decrease in the population of urban agglomerations as also newly declared towns with clear identifiable boundaries and landmarks. Each block has a population of 600–800 people, who form a part of approximately 200 households. Every year, 20% of the urban agglomerations are updated and over a period of 5 years, known as a 'phase', all blocks nationwide are updated. The use of up-to-date maps is, therefore, essential when replicating the NSS methodology for research.[3]

SECOND-STAGE SAMPLING

Selection of hamlet-groups and urban sub-blocks

Second-stage sampling is done in the field since no list of households is available. First, large villages are divided into hamlet-groups and large blocks into sub-blocks. The number of these subdivisions depends on the population of the FSU. The steps followed to delineate these are:

- Step 1: The approximate population of the village is ascertained by seeking the advice of the panchayat (village council), government school headmaster or zilla parishad (local government office).
- Step 2: The community groupings or 'hamlets' are identified and their geographical boundaries mapped.
- Step 3: The hamlets are grouped into NSSO 'hamlet-groups' for all villages with a population greater than 1200 (Table II).
- Step 4: Two hamlet-groups are randomly selected, using simple random sampling without replacement, and the listing of households is undertaken in these hamlet-groups.

For urban areas, roughly similar steps are taken for the identification of sub-blocks, but the sub-blocks comprise approximately 200 households. For the purpose of the NSS, two sub-blocks are selected using simple random sampling without replacement.[5]

Selection of households

To select the survey households, i.e. the second-stage strata (SSS) or ultimate sampling units (USU), households in the selected hamlet-groups and sub-blocks are listed. The methodology used is simple random sampling. The criteria for the selection of households are specific to the round being conducted. For example, in the healthcare and morbidity round the following criteria were used to select 10 households (Table III).

When these data are aggregated at a higher level, to avoid the sample from being purposive, representativeness is assured by multipliers which are to be used as weights. The NSSO provides these weights along with the unit-level data. These weights can be easily applied in statistical software such as SPSS or Stata. The data collected from these households are processed at the Kolkata division of the NSSO. These data can be purchased from the MOSPI.[5]

TABLE II. Criteria applied for dividing the village into hamlet-groups

Population	Number of hamlet-groups/sub-blocks formed
Less than 1200	None
1200–1799	3
1800–2399	4
2400–2999	5
3000–3599	6

Source: Reference 5

TABLE III. Criteria for selection of households in healthcare and morbidity round of National Sample Survey

Composition of substrata	With hamlet-group/ sub-block formation (for each)
SSS 1: households with at least one member hospitalized during last 12 months	4
SSS 2: from the remaining households, households with at least one child below the age of 5 years	2
SSS 3: from the remaining households, households with at least one member of the age of 60 years or above	2
SSS 4: other households	2
Sub-total	10

Source: Reference 5

Extraction of data in a usable format

The NSS data require extraction, tabulation and labelling, using appropriate software. The unit-level data are provided in a compact disc consisting of three folders. The first contains the data file 'in a flat line-sequential ASCII format,[6] which needs to be imported into a statistical software platform' (e.g. SPSS, MS EXCEL, Stata) for the purpose of analysis, since the raw data are provided as an array of numbers. The sequence of numbers reflects the questions in the survey questionnaire, which is divided into sections called 'levels'. The questionnaire for the healthcare and morbidity round had 16 levels, which covered information ranging from the household characteristics to the health profile of and healthcare utilization by individual householders.[5] Data pertaining to one or more states can be purchased.[6,7] The second folder includes household-level multipliers used to generate population-level estimates. This folder may be empty if the multipliers are provided in the data itself. The third folder includes two important supporting documents. One is a Microsoft Excel spreadsheet describing the layout of the data. The layout consists of the length of each variable and the position (known as byte position) determining the exact location of the variable in the array. It also provides the name of each variable which can be used to label the data. The second document is a 'read-me' file which explains how to use the layout file and how to generate multipliers and unique IDs to identify the unit of analysis (e.g. a household or individual).[6]

IMPLICATIONS FOR PUBLIC HEALTH

NSS data are among the most important sources of information for public health research in India. These data are highly validated. This is because first, they ensure that the fieldwork for the national

sample is carried out by teams of experienced and highly trained investigators from the Field Operations Division (FOD) of the NSSO. Second, each state conducts a parallel survey, supervised by the Directorate of Economics and Statistics. This survey uses a larger sample size and the sample overlaps with that of the nationally commissioned survey. Third, the FOD scrutinizes each questionnaire for logical correctness and the questionnaires found to be invalid are surveyed again. These steps ensure the validity of the survey.[3]

The survey on consumer expenditure has been repeated regularly and has provided valuable information on the trends in household expenditure, including expenditure on health (Appendix A, available at *www.nmji.in*).[8] Concerns regarding the high level of out-of-pocket expenditure on health, as revealed by the NSSs, have prompted policy-makers to plan for universal health coverage in India during the Twelfth Five-Year Plan.[9]

The data can also be used to assess the socioeconomic determinants of health and health-seeking behaviour. The healthcare and morbidity round, which is conducted on a decennial basis, is a rich source of data and includes information on hospital admissions, outpatient visits, healthcare expenditure (including means of financing healthcare), health-seeking behaviour, and the use of public and private healthcare facilities. In addition, the healthcare and morbidity round includes information on the socioeconomic profile of the household and was last conducted in 2004. These data make it possible to draw important comparisons within a state, as well as inter-state, national and international comparisons, and help to identify areas requiring specific action (Appendix B, available available at *www.nmji.in*).

In 2007, the Government of India, in collaboration with WHO, published a comparative analysis of public health indicators across the states, using three specific rounds of the NSSs (42nd, 52nd and 60th).[10] NSS data have also been particularly useful to researchers exploring patterns of health financing with a view to informing future policy on universal access to healthcare. A study published in 2008 highlighted the substantial financial burden of inpatient care for the uninsured population, particularly in rural areas, with the average out-of-pocket payments for hospitalization accounting for 11% and 9% of the total annual household expenditures in rural and urban areas, respectively.[11] A benefit incidence analysis published 3 years later by Acharya *et al*.,[12] using the data from the 52nd and 60th rounds to assess whether spending on public healthcare was becoming pro-poor, concluded that in Tamil Nadu, it had become much more pro-poor by 2004–05 than it had been a decade earlier, but that in Odisha, the benefits were limited to the outpatient services.[12]

NSS data can provide useful baselines for the evaluation of the impact of recently introduced health programmes that are targeted at specific population groups or have been launched in some states. In 2011, Fan *et al.* used the NSSO data from the 55th, 61st, and 64th rounds on consumer expenditure to examine variations in the roll-out of Andhra Pradesh's Rajiv Aarogyasri Scheme over time to evaluate the scheme's impact. They concluded that in its first phase, the scheme resulted in a 'significant reduction in out-of-pocket inpatient expenditure and, to a lesser extent, outpatient expenditure'. However, the scheme did not benefit the marginalized population as much as the rest of the population.[13]

The other areas in which public health researchers have utilized NSS data and drawn conclusions relevant for policy and practice have ranged from female labour[14] to disability.[15]

CHALLENGES IN USING THE NSSO DATA

The NSSO has an ethos of promoting and supporting research of major social relevance. Its reports and conclusions are easily accessible on public portals. Nevertheless, the utilization of the NSS data and methodology presents several challenges.

To replicate NSS sampling methods, urban blocks, as mapped by the NSSO, have to be used to make sure that the sample is representative and to ensure comparability, should the NSS data be used as the baseline. Furthermore, despite regular updating, the efforts to map urban blocks may not succeed in keeping pace with the rapidity of urbanization. As a result, the localities would not resemble the maps drawn recently, which would pose a challenge to the surveyor. Despite these limitations, the NSS urban blocks remain the best available resource for research including urban areas, as the Census wards may be even more outdated. At a practical level, the absence of a central repository for all maps is a source of inconvenience as maps have to be secured from the respective regional NSSO offices under which the geographical area covered by the research falls. In recognition of these concerns, the NSSO has digitized its recently updated maps, which are likely to be more easily accessible to researchers in the future.

The NSSs address specific national objectives and the collection of data is limited to achieving those objectives. For example, the data on health expenditure that are included in the consumer expenditure survey do not capture whether the healthcare providers are from the public or private sector, the diseases on which expenditure is incurred or the means of financing healthcare. Some of these data may be collected in the health and morbidity round, but it is not possible to link the data from different rounds of the NSS. This is because the surveys are conducted with different households.

CONCLUSION

The NSSs are a valuable source of information on India's population parameters and are of critical importance to policy-makers and researchers. They may have certain limitations, but they provide valuable insights into the health status of the general public and the trends across the country. Properly extracted and tabulated, NSS data can help a researcher in the analysis of trends and the evaluation of the impact of interventions. In addition, they are useful for drawing national and international comparisons of socioeconomic and population profiles. Thus, they can play an important role in influencing policies and interventions aimed at improving the health and well-being of the population.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support and encouragement of the International Development Research Centre, Canada, Wellcome Trust, the UK Department for International Development, and Rockefeller Foundation throughout the study. A special thanks to Adam Wagstaff of the World Bank and Rahul Ahluwalia of the Indian School of Business for mentoring us while we were learning to use the NSSO data. We also gratefully acknowledge the NSSO field officers who helped us understand the sampling technique.

REFERENCES

1 Mohanty B. Special issue on gender statistics. *Samikshya* [Internet]. 2011; 6(6). Available at *http://desorissa.nic.in/des-data/latest_publications/samikshya-2011.pdf* (accessed on 1 Dec 2012).
2 Som RK. *A manual of sampling techniques.* 1st ed. London:Heinemann; 1973.
3 Ministry of Statistics and Programme Implementation. *Concepts and definitions used in NSS [Internet].* New Delhi:MOSPI; 2001. Available at *http://mospi.nic.in/mospi_new/upload/nsso/concepts_golden.pdf?status=1&menu_id=49* (accessed on 1 Dec 2012).
4 Ministry of Statistics and Programme Implementation. Microdata archive—Data Catalog [Internet], 2012. Available at *http://164.100.34.58/index.php/catalog* (accessed on 1 Dec 2012).
5 National Sample Survey Organization. *Morbidity, healthcare and condition of the aged report. Report No. 507 (60/25.0/1).* New Delhi:Ministry of Statistics and

Programme Implementation, 2004. Available at *http://mospi.nic.in/rept%20_%20pubn/507_final.pdf* (accessed on 3 May 2012).
6 Ministry of Statistics and Programme Implementation. Government of India, Ministry of Statistics and Programme Implementation (Computer Centre) rate list of unit-level NSS data on CD's [Internet]. New Delhi:MOSPI; 2012. Available at *http://mospi.nic.in/mospi_new/upload/nsso/ratelist_UnitData.pdf* (accessed on 1 Dec 2012).
7 MOSPI. Ministry of Statistics and Programme Implementation official website [Internet], 2012. Available at *http://mospi.nic.in/Mospi_New/site/inner.aspx?status=3&menu_id=31* (accessed on 29 May 2012).
8 National Sample Survey Organisation. India—Household Consumer Expenditure in 2006–07. NSS 63rd round, June 2006–July 2007. New Delhi:NSSO, Ministry of Statistics and Programme Implementation; Report No. 527 (63/1.0/1). 2004. Available at *http://164.100.34.62/index.php/catalog/43* (accessed on 19 Feb 2013).
9 Bang A, Chaterjee M, Dasgupta J. *High Level Expert Group Report on Universal Health Coverage for India. Instituted by the Planning Commission of India.* New Delhi:Planning Commission of India; 2011. Available at *http://planningcommission.nic.in/reports/genrep/rep_uhc0812.pdf* (accessed on 1 Dec 2012).
10 Ministry of Health and Family Welfare. *Select Health Parameters: A comparative analysis across the National Sample Survey Organization (NSSO). 42nd, 52nd and 60th rounds.* New Delhi:MoHFW and WHO; 2007. Available at *http://s3.amazonaws.com/zanran_storage/whoindia.org/ContentPages/112354873.pdf* (accessed on 1 Dec 2012).
11 Flores G, Krishnakumar J, O'Donnell O, van Doorslaer E. Coping with health-care costs: Implications for the measurement of catastrophic expenditures and poverty. *Health Econ* 2008;**17:**1393–412.
12 Acharya D, Vaidyanathan G, Muraleedharan V, Dheenadayalan D, Dash U. *Do the poor benefit from public spending on healthcare in India? Results from benefit (utilisation) incidence analysis in Tamil Nadu and Orissa.* London:The Consortium for Research on Equitable Health Systems 201; pp 1–38. Available at *http://www.dfid.gov.uk/r4d/PDF/Outputs/Equitable_RPC/India_BUIA.pdf* (accessed on 30 Nov 2012).
13 Fan V, Mahal A, Karan A. State health insurance and out-of-pocket health expenditures in Andhra Pradesh, India. Centre for Global Development Working Paper No. 298, 2012; pp 1–53. Available at *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2102772##* (accessed on 1 Dec 2012).
14 Shanthi K. *Female labour migration in India: Insights from NSSO Data Female Labour Migration in India: Insights from NSSO Data. Working Paper no. 4/2006.* Chennai:Madras School of Economics; 2006: 1–53. Available at *http://www.mse.ac.in/pub/santhi_wp.pdf* (accessed on 24 Jul 2012).
15 Walia GK. *Disability in India.* New Delhi:South Asia Network for Chronic Disease; 2011. Available at *http://sancd.org/uploads/pdf/disability.pdf* (accessed on 29 Aug 2012).

Introducing Images/Videos

These should be interesting or unusual educational images related to healthcare with a short description (up to 250 words). Examples include radiological or microscopic images or pictures of a pathological entity, a clinical sign or even a phenomenon. Anything of interest to readers of the Journal may be submitted. The image/video should either belong to the author or be accompanied by a letter of permission from the photographer. Images/videos may be in black and white or in colour, and should be in JPG or TIFF format/MPEG2 or MPEG4 video format. A video could be stand-alone or complement a still image (e.g. for an echocardiographic recording, an endoscopic recording, a neurological sign). Videos will be placed on the website of the Journal with a description and web address (URL) in the print version of the Journal. All images that allow for possible identification of the patient must be accompanied by a clear consent from the patient to publish the image.