Data Quality of Select Economic Indicators in a National Survey Using Handheld Device: Experience from Global Adult Tobacco Survey, India (2009-2010)

Pratap K. Jena1,2,*, Sudeep K. Shetty3, Jugal Kishore1, Monali Swain4
1Department of Public Health, KSHEMA, Nitte University, Mangalore, India
2Health System Research Initiative India, Bangalore, India
3Department of Community Medicine, Maulana Azad Medical College, New Delhi, India
4Tata Memorial Hospital, Mumbai, Maharashtra, India

Received Date: June 08, 2015, Accepted Date: October 13, 2015, Published Date: October 23, 2015.
*Corresponding author: Pratap K. Jena, Department of Public Health, KSHEMA, Nitte University, Deralakatte, Mangalore-575018, India, Tel: +91-812-317-7278; Fax: 0824-220-4162; E-mail: drpratapjena@gmail.com

Abstract

Introduction: Global Adult Tobacco Survey (GATS)-India is the first national survey in which electronic handheld devices were used for data collection and data management. Apart from the participant selection during household survey, rapid data collection, data collation, cost effectiveness, acceptability and improved data quality are some of the key benefits of handheld device use. However, the data quality of GATS-India survey for certain variables like cigarette and rolled cigarette use, quit attempt and duration of abstinence has been questioned in the recent past. In this context, this study is undertaken to understand the usefulness of handheld device in data collection during GATS-India (2009-10) survey, considering data quality of select economic indicators of tobacco use.

Methods: Economic indicators of tobacco use (last purchase and the price paid for it) from items F01 and F02 of GATS-India tool was analyzed to assess the quality of data using cross-tabulation, correlation, ‘t’ test and internal consistency test. Due to price variation from state to state and brand to brand, state stratified internal consistency analysis was done after limiting analysis to a single brand of manufactured cigarette which was most commonly reported in the survey.

Results: Data analysis suggests 90% of current manufactured cigarette users had purchased cigarettes for their use and 8.7% of daily users had never purchased any manufactured cigarette for their use. The number of the manufactured cigarette purchased by current users was varied between one and 179,820 units. The amount paid per cigarette stick was ranged from nil to 1,944 INR (1 USD = 65.5 INR). Around 9% packs-purchasing smokers and 83% carton-purchasing smokers poorly reported regarding the number of cigarettes per pack/carton. Less than daily smokers spent a higher amount than daily smokers (7.7 vs. 5.1 INR) during their last purchase. Despite of positive and very weak correlation for the entire sample, internal consistency between quantity of cigarette purchased and the price paid was either poor or unacceptable or unreliable in 11 states.

Conclusion: Handheld electronic device has been successfully used in developed nations and has an equal potential to be used in India. However inconsistent responses for selected economic indicators, highlight limited validation of the GATS-India tool, which could underestimate its usefulness. Full utilization of automated features available in handheld devices could reduce inconsistent responses in future surveys.

Keywords: Handheld device; PDA; Population Survey; GATS; India; Date quality

Introduction

India has a population of approximately 1.21 billion distributed across 640 districts covering 35 States/Union Territories [1]. This population is heterogeneous in many aspects. Therefore, the number of interviews conducted in national surveys like National Family Health Surveys (NFHS) and District Level Household Surveys (DLHSSs), can be as high as 0.7 million [2,3]. Considering this high number of interviews, usual pen-and-paper based data collection; doubled data entry and rechecking consume time and other resources along with possessing higher risk of poor data quality arising out of human error. Keeping this in mind, electronic handheld devices were used for data collection and data management in Global Adult Tobacco Survey (GATS), India (2009-10) [4].

Use of handheld devices or the personal digital assistant (PDA) for data collection in resource poor settings could simplify the process of questionnaire development and data collection [5]. Inclement weather condition and multiple language settings have limited effect on data collection while using PDAs. Automated features in PDAs can help skipping inappropriate items easily and also could highlight error for a given input with suggestions for correction. Thus, rapid and reliable data collection is possible while using PDAs. Limitations like battery backup, input errors, requirement of computer-literate users and IT infrastructure, data storage, loss of data and higher startup costs, during use of PDAs are of concern [5]. However Kizito et al [6] neither encountered any PDA-related issues nor data loss, in his study that had collected data from 83,346 persons over a period of seven-week. This study had demonstrated higher acceptability, completeness of survey data, and time efficiency. Similar observation along with increased data quality was noted in other studies [7-11].

As GATS-India had used handheld devices in the survey, it is normal to expect higher data quality. However a study by Giovino et al [12] on 14 countries GATS data, indicated a peculiar anomaly suggesting female daily smokers use more number of cigarettes per day than their male counterparts in India (not in other 13 countries) only. This anomaly was questioned in the pretext of poor data quality and other issues in various studies [13-17]. This study is further expanded to examine data quality while assessing economic indicators of tobacco use (last purchase and price paid for it) from items F01 and F02 of GATS-India survey tool. Other indicators of usefulness of handheld device as described as above are beyond the scope of this paper and described elsewhere.

Methods

The GATS-India questionnaire pertaining to the number of cigarette purchased last time (F01) and price paid for it (F02) among current cigarette smokers were critically examined using GATS syntax (GTSS 2011), GATS-India code book (2011) and GATS-India report (2010) and [4,18,19]. Appropriate logic to validate
responses was ascertained from these books with the help of experts. Current (daily and less than daily) manufactured cigarette smokers were assessed for their invalid responses during reporting about their last purchase (quantity and price) of cigarette in the section F (items F01 and F02 respectively) of the GATS-India tool.

Item F01 required two distinct steps to be followed to obtain the required data [19]. Step one involved mentioning the number of unit of cigarette purchased, i.e. loose cigarettes, packs, cartons or others (specification required). In second step, number of cigarettes per unit of purchase need to be mentioned. These two steps could help the analyst to estimate the number of manufactured cigarette purchased last time. In India, one manufactured cigarette pack contains either 10 or 20 number of individual cigarettes and a carton contains 10 packs, i.e. 100-200 cigarettes. Other form of ‘unit’ of manufactured cigarette sale or purchase is not known. These facts were considered while critically examining the F01 & F02 item.

These items from the GATS-India data available in public domain were analyzed using IBM SPSS (v.20) to get first-hand information regarding data quality in GATS-India. To identify invalid responses, related items were cross tabulated. Correlation between number of cigarettes purchased and price paid for it were compared using correlational statistics. Mean number of cigarettes purchased by ‘daily’ and ‘less than daily’ users were compared using independent ‘t’ test. A sub-analysis for internal consistency (to estimate Cronbach’s Alpha) assessment was made. The analysis was also limited to a single brand (the highest reported) to avoid potential misinterpretation arising out of differential pricing of cigarettes across states and brands. GATS sample weight was not used considering random sub-sample analysis in this study.

**Results**

Analysis suggests 4893(90%) current cigarette smokers (daily and less than daily) had reported about the purchasing of cigarettes/packs/cartons/other units, during their last purchase for their own use. When their smoking status examined, it was found that around 8.7% of daily cigarette smokers and 8.6% of less than daily smokers never bought cigarette for their use as detailed in the Table 1.

The range of ‘units of cigarette’ purchased by the current users was ranged between one and 999 (cigarettes: 1 - 999, packs: 1 - 999, cartons: 1 - 444). About 1735 smokers reported that they had bought cigarette packs and among them 80.2% mentioned 10 cigarettes per pack, 11.1% reported 20 cigarettes per pack and rest mentioned other numbers ranging from one to 200. Total 29 smokers reported of buying cartons during last purchase and the rest mentioned other numbers ranging from one to 200. The range of units of cigarettes purchased per unit of purchase need to be mentioned. These two steps could help the analyst to estimate the number of manufactured cigarette purchased last time. In India, one manufactured cigarette pack contains either 10 or 20 number of individual cigarettes and a carton contains 10 packs, i.e. 100-200 cigarettes. Other form of ‘unit’ of manufactured cigarette sale or purchase is not known. These facts were considered while critically examining the F01 & F02 item.

When quantity of cigarettes bought was estimated as per the GATS syntax guideline, it was found that the number of cigarettes bought during last purchase ranged between one and 179,820.

**Table 1: Cross tabulation of item B01 (Smoking status) and Unit of Last Purchase of cigarette.**

| F01a. The last time you bought cigarettes for yourself, how many cigarettes did you buy? [ENTER UNIT] | B01. Do you 'currently' smoke tobacco on a daily basis, less than daily, or not at all? |
|---|---|---|
| CIGARETTES | DAILY | LESS THAN DAILY | NOT AT ALL |
| 2000 (51.8%) | 1109 (74.1%) | 0 |
| PACKS | 1497 (38.8%) | 237 (15.8%) | 1 (100%) |
| CARTONS | 16 (0.4%) | 13 (0.9%) | 0 |
| OTHER SPECIFY | 12 (0.3%) | 8 (0.5%) | 0 |
| NEVER BOUGHT CIGARETTES | 334 (8.7%) | 129 (8.6%) | 0 |

Frequency table of F02 suggests the price paid during last purchase was between nil to 7,777 INR (for the highest sold single brand only, 65.5 INR = 1 USD) with some people being not aware of the price they paid for purchasing cigarette. When the price paid per each manufacturing cigarette purchased was estimated, it was ranged between nil to 1,944 INR with 50%(1) paying less than one INR. Considering the mean amount paid for their last cigarette purchase, less than daily users (7.7 INR) had spent more money than daily users (5.1 INR). However, the difference in mean was not significant during the independent student ‘t’ test.

Very weak positive correlation (r = 0.102; p < 0.001) was observed between the number of cigarettes purchased and price paid when the entire data set was analyzed. In state specific analysis, three states indicated negative correlation. During an internal consistency check for the quantity of cigarette purchased and price paid, Cronbach’s Alpha value was good in 12 states, acceptable in eight states, poor/unacceptable in 8 states and not reliable in three states as detailed in the Table 2.

**Discussions**

The data quality in this self-reported survey would depend on respondents, interviewer, and the tools used. GATS-India has used a standard tool which was programmed into PDAs to create a data collection interface. The data thus collected was made available in the public domain after data mining. However, in this study non-purchasing of cigarette by smokers for their use, reporting of invalid units when asked about number cigarettes per pack (other than 10 or 20) or cartons (other than 100 or 200), failure to specify the ‘other’ unit of cigarette purchase, purchase by never users for their own use, purchasing of 999 packs or 405 cartons, nil paid amount for purchase, cigarette price less than one INR, purchasing of 179,820 cigarettes for use, higher average price paid by ‘less than daily’ users, negative correlation and poor internal consistency between the numbers of cigarettes purchased and price paid etc. highlight the poor quality of data.

The self-reported responses may be true and PDAs might have captured the same. For example, a new initiator of smoking might be borrowing cigarette from friends, and therefore it could be possible that they never bought cigarettes for their use. While such logic is justifiable for less than daily smokers, it may not hold good for the daily smokers. Another important aspect is, the retail credit system is very common in India, where people buy from a local retail shop on credit and pay them later. Such retail credit may be for the entire amount or part of it. Therefore, it is possible that the response against the price paid for last purchase could be nil or less than the actual price of cigarette unit purchased. Since the programmed tool has not considered the same, data on actual price for the last purchase of cigarettes might be misleading.

Following reporting of 14 country GATS data by Giovino et al [12], a correspondence [13] also published, contending the gender-reversed finding about cigarette smoked per day in India. The authors attributed such findings to definitional issues, poor piloting weak survey technique and misreporting. The authors...
had also pointed out that smokeless tobacco product information had been collected under other smoking product use and thereby questioned internal validity of the GATS-India survey data. Giovino et al [20] in their reply avoided the actual contention and argued that manufactured cigarettes and hand-rolled cigarettes need to be separated from bid is owing to their differential sources, marketing and users profile, but at the same time they failed to advance similar argument [14] as the basis of product specific separate analysis of manufactured cigarette and rolled-cigarette. However, recording of smokeless tobacco product as well as duplicate records of classified smoking product in the other smoking category is clearly evident, suggesting limited validation of data collecting interface [14]. Jena et al [17] further elaborated the concept of interaction between different nicotine containing tobacco products that have led to Simpson’s Paradox . Apart from these, high digit bias (modified Whipple Index - 226.3) while collecting information on the quantity of cigarette purchased per day among daily users, which further reconfirms poor data quality in GATS-India survey [15]. The authors have also identified the imprecise definition of ‘current daily user’ and ‘current less than daily use’ resulting in more than one-fifth invalid responses regarding duration of abstinence [16]. In these studies authors had expressed concern over lost opportunity to internally validate responses in the GATS-India data collecting tool. The authors also had suggested input errors considering the internal validity of the items could have yielded better quality data.

The poor quality of data as highlighted in this study per-se is not a fault of using PDAs. As identified in earlier studies, lack of internal validation of the tool including poor programming of PDAs due to human error is the main determinant of such results. For example, purchasing of cigarette by never users could have been highlighted as an error in data input during data collection leading to further clarification and correction. Similarly, reporting of number of cigarettes per pack other than 10 or 20, and reporting of number of cigarettes per carton other than 100 or 200 could have been easily identified by the well-programmed handheld device and thus could have reduced the presence of inconsistent or invalid responses to a great extent. If the automated features were fully utilized while programming of PDAs, many errors like nil paid amount for the purchase, cigarette price less than one INR, purchasing of 179,820 cigarettes for use, etc. could have been easily rectified during data collection itself.

The GATS report (2010) suggests that use of handheld device helped in the random selection of individuals for interview during household survey, facilitated ‘skip patterns’ to fill the GATS India questionnaire, and improved data quality through some built-in validity checks during the process of data collection and faster data collation. The entire process of survey questionnaire administration, data collection and collation using the handheld device was pre-tested. Two specific manuals, ‘Handheld (IPAQ) Manual’ and ‘Data Transfer Manual’ were developed to standardize the process. However such claims are partially true.

The fieldwork for GATS took place from June 2009 to January 2010. The report was published in September 2010. It took seven months to be in the public domain, which may be attributable to various administrative issues of project management. While time efficiency in generating report was not demonstrated in this survey, there is no doubt that the use of hand held device could improve time efficiency of a given survey.

**Conclusion**

The handheld electronic device has been successfully used in developed nations and has an equal potential to be used in India. However inconsistent responses for select economic indicators, highlight limited utilization of automated features available in handheld devices to check internal consistency of respondents. The poor quality of data as highlighted in this study per-se is not a fault of using PDAs in the survey. Validation of digital data collection interface could reduce inconsistent responses to a great extent in future surveys.

**References**

1. Office of the Registrar General & Census Commissioner, India. Census India, 2011. [Cited 2014 Dec 4]. Available from: http://www.censusindia.gov.in/.
2. International Institute for Population Sciences (IIPS). National Family Health Survey (NFHS): Reproduction and Child Health Project. 2014 [Cited 2014 Dec 14] Available at http://www.rchiips.org/NFHS/index.shtml.
3. International Institute for Population Sciences (IIPS). District Level Household and Facility Survey (DLHS): Reproductive and Child Health Project. 2014 [Cited 2014 Dec 14] Available at http://www.rchiips.org/.
4. Ministry of Health and Family Welfare (MoHFW), Govt. of India. The Global Adult Tobacco Survey (GATS) India (2009-2010). New Delhi: Ministry of HPW, Govt. of India 2010.
5. Seebregts CJ, Zwerveren M, Mathews C, Fairall L, Fisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in poor-resource settings: development and evaluation of PDACt, a Palm Pilot interviewing system. Int J Med Inform. 2009;78(11):721-31. doi: 10.1016/j.ijmedinf.2008.10.006.

6. Shirima K, Mukasa O, Schellenberg JA, Manzi F, John D, Mushy A, Mrisho M, et al. The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania. Emerg Themes Epidemiol. 2007;4:5.

7. Tegang SP, Emukule G, Wambugu S, Kabom I, Mwarogo P. A comparison of paper-based questionnaires with PDA for behavioural surveys in Africa: findings from a behavioral monitoring survey in Kenya. Journal of Health Informatics in Developing Countries 3: 22-25.

8. Marchant T, Schellenberg J, Peterson S, Manzi F, Waiswa P, Hanson C, Temu S, et al. The use of continuous surveys to generate and continuously report high quality timely maternal and newborn health data at the district level in Tanzania and Uganda. Implement Sci. 2014;9:12. doi: 10.1186/s13012-014-0112-1.

9. Zhou Y, Lobo NF, Wolkon A, Gimnig JE, Malishee A, Stevenson J, et al. PGMS: a case study of collecting PDA-based geo-tagged malaria-related survey data. Am J Trop Med Hyg. 2014;91(3):496-508. doi: 10.4269/ajtmh.13-0652.

10. Yu P, de Courten M, Pan E, Galea G, Pryor J. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. Int J Med Inform. 2009;78(8):532-42. doi: 10.1016/j.ijmedinf.2009.03.002.

11. Bugembe R, Watareashne C, Dinwo J, Mundida O, Langhaug L, Willis N, et al. Positive attitudes to pediatric HIV testing; findings from a nationally representative survey from Zimbabwe. PLoS One. 2012;7(12):e5213. doi: 10.1371/journal.pone.005213.

12. Giovino GA, Mirza SA, Samet JM, Gupta PC, Jarvis MJ, Bhala N, et al. Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. Lancet. 2012;380(9842):668-79. doi: 10.1016/S0140-6736(12)61085-X.

13. Jena PK, Kishore J, Bandyopadhyay C. Prevalence and patterns of tobacco use in Asia. Lancet. 2012;380(9857):1906; author reply 1906-7. doi: 10.1016/S0140-6736(12)62108-4.

14. Jena PK, Kishore J, Sarkar BK. Global Adult Tobacco Survey: A case for change in definition, analysis and interpretation of “Cigarette” and “Cigarettes per Day” in completed and future surveys. Asian Pac J Cancer Prev. 2013;14(5):3299-304.

15. Jena PK, Kishore J, Jahnavig. Correlates of Digit Bias in Self-reporting of Cigarette per Day (CPD) Frequency: Results from Global Adult Tobacco Survey (GATS), India and its Implications. Asian Pac J Cancer Prev. 2013;14(6):3865-9.

16. Jena PK, Kishore J, Patt S, Sarkar BK, Das S. Tobacco Use and Quit Behaviour Assessment in the Global Adult Tobacco Survey (GATS): Invalid Responses and its implication. Asian Pac J Cancer Prev. 2014;14(11):6563-8.

17. Jena P, Sudeep K, Kishore J (2014) Global Adult Tobacco Survey, India (2009-10): Why There is Need for Improvement? Proceedings of International Conference of Asian Pacific Organization for Cancer Prevention.

18. Global Tobacco Surveillance System (GTSS) Group. GTSS: Global Adult Tobacco Survey (GATS): Indicator Guidelines: Definition and Syntax. 2009.

19. Global Tobacco Surveillance System (GTSS) Group. 2011. GTSS: Global Adult Tobacco Survey: India 2009 Codebook.

20. Giovino GA, Gupta PC, Samet JM. Prevalence and patterns of tobacco use in Asia – Authors’ reply. Lancet. 2012;380(9857):1906–07.