

RESEARCH ARTICLE

Correlates of Digit Bias in Self-reporting of Cigarette per Day (CPD) Frequency: Results from Global Adult Tobacco Survey (GATS), India and its Implications

Pratap Kumar Jena^{1*}, Jugal Kishore², G Jahnavi³

Abstract

Background: Cigarette per day (CPD) use is a key smoking behaviour indicator. It reflects smoking intensity which is directly proportional to the occurrence of tobacco induced cancers. Self reported CPD assessment in surveys may suffer from digit bias and under reporting. Estimates from such surveys could influence the policy decision for tobacco control efforts. In this context, this study aimed at identifying underlying factors of digit bias and its implications for Global Adult Tobacco Surveillance. **Materials or Methods:** Daily manufactured cigarette users CPD frequencies from Global Adult Tobacco Survey (GATS) - India data were analyzed. Adapted Whipple Index was estimated to assess digit bias and data quality of reported CPD frequency. Digit bias was quantified by considering reporting of '0' or '5' as the terminal digits in the CPD frequency. The factors influencing it were identified by bivariate and logistic regression analysis. **Results:** The mean and mode of CPD frequency was 6.7 and 10 respectively. Around 14.5%, 15.1% and 15.2% of daily smokers had reported their CPD frequency as 2, 5 and 10 respectively. Modified Whipple index was estimated to be 226.3 indicating poor data quality. Digit bias was observed in 38% of the daily smokers. Heavy smoking, urban residence, North, South, North-East region of India, less than primary, secondary or higher educated and fourth asset index quintile group were significantly associated with digit bias. **Discussion:** The present study highlighted poor quality of CPD frequency data in the GATS-India survey and need for its improvement. Modeling of digit preference and smoothing of the CPD frequency data is required to improve quality of data. Marketing of 10 cigarette sticks per pack may influence CPD frequency reporting, but this needs further examination. Exploring alternative methods to reduce digit bias in cross sectional surveys should be given priority

Keywords: Smoking - cigarette per day (CPD) - digit bias - smoothing - GATS - India

Asian Pacific J Cancer Prev, 14 (6), 3865-3869

Introduction

Digit bias has been widely discussed in tobacco epidemiological survey for self-reporting of cigarette per day (CPD) frequency (Means et al., 1992; Klesges et al., 1995; IARC, 2008). Digit bias may occur due to rounding to multiples of a base unit (a convenient and significant number i.e. 5, 10 etc.) or due to terminal digit preference (or avoidance), resulting in data 'heaping' (Crocketta et al., 2001; IARC, 2008). The rounding of digits appears to round down more often than round up leading to underreporting of CPD use in tobacco surveys (Warner et al., 1978; Hatziandreu et al., 1989).

CPD is an important measure of smoking behaviour and is a key indicator for Global Adult Tobacco Surveillance (Giovino et al., 2012). Surveys are valid ways to estimate tobacco use prevalence (Caraballo et al., 2001; 2004) but the accuracy of CPD data may suffer from digit bias

like reporting of excess frequencies at round or preferred figures (Klesges et al., 1995; IARC, 2008). Digit bias in CPD frequency reporting could represent data quality and consistency (Means et al., 1992).

CPD measure is an integral part of all tobacco induced cancer mortality and morbidity (dose response) studies (Law et al., 1997; IARC, 2004). Cigarette per day is a classical measure of nicotine dependence and its severity (Dawe et al., 2002). Nicotine dependence assessment tools assess CPD frequency interval (<10, 11-20, 21-30 and >30) to denote severity of nicotine dependence (Heatherton et al., 1991; Chaiton et al., 2007). Digit bias would result in misclassification of high dependence as low dependence and could lead to suboptimal cessation therapy (Fagerstrom, 2003). Further the decision on treatment options is also dependent on frequency of CPD reported. For example nicotine replacement therapy is more suitable for those smokers who smokes ≥ 10 cigarette

¹Research, Public Health Foundation of India, ²Department of Community Medicine, Maulana Azad Medical College, New Delhi, ³Department of Community Medicine, Madha Medical College or Research Institute, Chennai, India *For correspondence: drpratapjena@gmail.com

per day (West et al., 2000). The 2 mg nicotine gum is most suitable for smokers using <25 CPD and the 4 mg gum is most appropriate for smokers using ≥25 CPD (USDHHS, 2008; DGHS, 2011). Therefore rounding may affect medication prescription for smokers. In view of all these implications, more accurate measurement of CPD use is desirable.

The CPD use is being measured by subjective global self report, timeline follow-back (TLFB) method and real time ecological momentary assessment (EMA) (Shiffman, 2009; Berkman et al., 2011). Former two methods show four to six time higher incidence of digit bias than latter one (Shiffman et al., 2009). Also prospective CPD data collection (Perkins et al., 2012) has twice less digit bias than retrospective CPD data collection technique. These methods are time consuming and resource intensive.

India is home to 111 million smokers (IIPS, 2010). There is no information on the self reported CPD data quality in India. Therefore, it is pertinent to understand the CPD use data quality, magnitude of digit bias and factors influencing it. This study is aimed at quantifying and identifying correlates of digit bias while reporting CPD frequency in the GATS-India survey and to discuss its implication.

Materials and Methods

Global Adult Tobacco Survey (GATS) is nationally representative survey and was conducted in 2009-10 in India (IIPS, 2010). Item no (06a) in GATS questionnaire (On an average, how many manufactured cigarettes do you currently smoke each day?) was asked to daily smokers to get CPD frequency data.

For the purpose of this study we analyzed GATS-India data available in the public domain from the CDC website (CDC web data, 2012). A subsample of 3411 daily manufactured cigarette users' data was analyzed. The CPD frequency with last digit as '0' or '5' was assumed as digit biased. Hence CPD frequency was classified as digit biased or not digit biased. Bivariate analysis (using chi-square test) and binomial regression analysis was done to identify socio-demographic correlates of digit bias in CPD reporting. There is wide cultural variation in different regions of India. Therefore 'National regions' was used as proxy to represent such cultural variations. Principal component analysis was carried out to calculate asset index quintile groups. Considering low mean number of cigarette use in Indian population (Giovino et al., 2012), CPD use as ≥15 was defined as heavy user. As analyses were conducted on a non-random subsample, variance and sampling weights were not used.

Modified Whipple Index (WI) was calculated to assess heaping of terminal digits '0' and '5' in reporting of CPD. WI modification procedure adopted by Wang et al. (1995), Danic et al. (2004) and Shiffman et al. (2009) were followed to adapt it for validating CPD frequency data quality reported in the GATS survey. Adapted WI formula is given below for CPD frequency distribution between 3 and 17 considering occurrence of digit preference in multiples of 5. We have limited CPD frequency analysis for WI calculation, as 94% respondents reported the CPD

frequency ≤17.

$$WI = [100 \times \Sigma(P5 + P10 + P15)] / [0.2 \times \Sigma(P3 + P4 + P5 + \dots + P17)]$$

P is the size of population (number of respondents) of a particular CPD frequency. Whipple's index varies between 0 and 500. A value of 0 indicates that digits '0' and '5' are not reported, 100 means there is no preference for '0' or '5', and 500 means only the digits '0' and '5' are being reported in the frequency data. WI inferences as adopted by above authors are as follows: <105=highly accurate; 105-109.9=fairly accurate; 110-124.9=approximate; 125-174.9=rough; ≥175=very rough (Perdeshi, 2010).

Results

In the GATS-India survey 3411 respondents reported daily cigarette use in between 1 and 110 with mean CPD as 6.7. The self reported CPD frequency among daily manufactured cigarette users had been represented in Figure 1. About 44.2% of the daily smokers reported CPD frequency as <5. The mode of CPD frequency distribution was 10 followed by 5 and 2. The CPD frequency as 2, 5, 10 was reported by 14.5%, 15.1% and 15.2% respectively. The figure clearly indicates heaping of CPD frequency around terminal digit '5' or '0'. Also heaping was observed around even number digits like 2, 8, 12 digits. The WI was calculated to be 226.28 (95% CI: 216.45-236.10). Overall 38% (95% CI: 36.34-39.59%) of these respondents had shown digit bias. However, when CPD frequency <5 excluded and analyzed the digit bias increased to 67.9%.

Bivariate analysis (Table 1) indicated that male, urban and heavy user reported more digit bias than their respective counterparts which was found to be highly significant (P≤0.001). Among the national regions, respondents in the South and North East regions had shown higher digit bias than other regions. The digit bias was significantly different among the regions. Among the occupational groups, employed and students reported significantly more digit bias than other groups. There was clear and significant increase in digit bias with increase in educational level and asset index quintile groups was observed. Similar trend was observed in case of age but digit bias reporting was reduced after the age of 54 years.

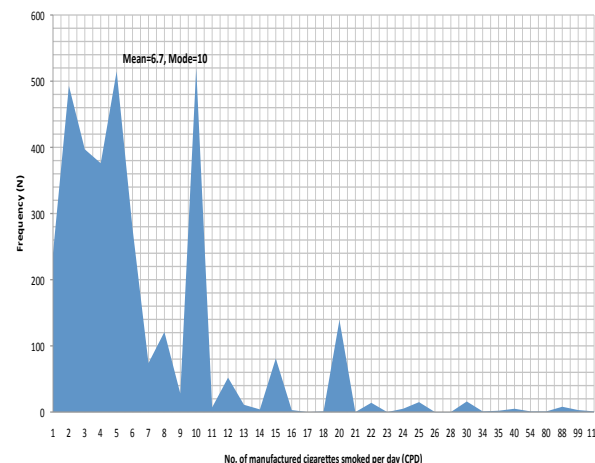


Figure 1. Self-Reported Cigarettes Per Day Frequency Data Showing Heaping

Table 1. Digit Bias in Self Reporting of Frequency of Cigarette Per Day (CPD)

		Not Digit Biased	Digit Biased	P value	Adj. OR	95% CI	P Value
Overall		2116(62.0)	1295(38.0)	0.001	-		
Type of residence	Rural	1219(64.5)	670(35.5)		Reference Category		
	Urban	897(58.9)	625(41.1)		1.40	1.15-1.70	0.001
National region	Central	161(75.6)	52(24.4)	<0.001	Reference Category		
	North	390(63.3)	226(36.7)		1.88	1.19-2.97	0.007
	South	347(56.9)	263(43.1)		1.99	1.28-3.09	0.002
	East	281(67.5)	135(32.5)		1.06	0.65-1.72	0.817
	North East	802(59.0)	557(41.0)		1.81	1.19-2.76	0.006
	West	135(68.5)	62(31.5)		1.16	0.67-2.02	0.587
Gender	Male	1979(61.3)	1248(38.7)	<0.001	Reference Category		
	Female	137(74.5)	47(25.5)		0.89	0.55-1.43	0.625
Education Level	No formal education	362(70.7)	150(29.3)	<0.001	Reference Category		
	less than primary	224(62.4)	135(37.6)		1.40	1.03-1.92	0.033
	Primary but less than Secondary	704(62.6)	420(37.4)		1.22	0.93-1.60	0.15
	Secondary or above	286(56.2)	223(43.8)		1.48	1.08-2.03	0.015
Occupation Status	Employed	802(59.5)	546(40.5)	0.003	Reference Category		
	Student	949(61.9)	584(38.1)		1.03	0.84-1.25	0.796
	Self Employed	80(69.6)	35(30.4)		0.7	0.33-1.49	0.353
	Other*	169(65.3)	90(34.7)		0.86	0.59-1.24	0.413
	Home Maker	111(73.5)	40(26.5)		0.87	0.54-1.40	0.557
Asset Index Quintile Group	Lowest	267(73.2)	98(26.8)	<0.001	Reference Category		
	Second	434(66.2)	222(33.8)		1.2	0.87-1.64	0.268
	Middle	386(63.2)	225(36.8)		1.1	0.79-1.54	0.559
	Fourth	527(57.9)	383(42.1)		1.47	1.06-2.04	0.02
	Highest	502(57.8)	367(42.2)		1.22	0.83-1.78	0.306
Age Group	15-24 Yrs	217(65.4)	115(34.6)	0.064	Reference Category		
	25-34 Yrs	548(64.2)	306(35.8)		1.16	0.81-1.67	0.426
	35-44 Yrs	633(60.6)	412(39.4)		1.28	0.90-1.83	0.173
	45-54 Yrs	412(58.5)	292(41.5)		1.34	0.92-1.95	0.122
	55 Yrs Plus	306(64.3)	170(35.7)		1.27	0.85-1.89	0.251
Heavy User	No	2080(66.8)	1035(33.2)	<0.001	Reference Category		
	Yes	36(12.2)	260(87.8)		14.49	9.69-21.66	<.001

*Other means 'retired or unemployed'

However in the regression model, age, gender, occupation turned insignificant predictor of digit bias reporting. The model indicates that being a heavy user, urban resident, belonging to North/South/North East region of India, less than primary or secondary/above educated; and fourth asset index quintile group were significant predictors of digit bias reporting than non heavy users, rural resident, belonging to Central region of India, not formally educated and lowest asset quintile group respectively.

Discussion

Study results indicate that there is digit bias ('0' or '5') while reporting CPD frequency in the GATS India survey. About two in five daily smokers report digit bias with 10 as the most commonly reported CPD frequency. Whipple Index suggests quality of CPD frequency data of GATS-India survey is very rough with heaping at frequencies having terminal digits as '0' or '5'. Relation of heavy use, urban residence and national region with digit bias is well explainable.

The digit bias reported here indicates that self reported CPD use data is less accurate. Earlier Klesges et al. (1995) had observed 71% digit bias among the respondents and found higher odds of digit bias in CPD reporting among the heavy (CPD>20) user than light one. Same study also indicated duration of education (OR: 0.95) and being

African or American (OR: 0.6) has protective role in digit bias while reporting CPD. In this study, education has no protective role in reporting biased CPD frequency, which needs further investigation. As 44.2% of daily cigarette smokers in this study reported CPD frequency less than five, the current definition, limited the probability of occurrence of higher digit bias. However, when smoker with CPD frequency under five is excluded, digit biased CPD frequency reporting increased to 67.9%. As two is the third most frequently reported figure for CPD use, digit preference other than '0' or '5' may need to be explored.

The self-report digit "spikes" or "heaping" at particular pattern (multiples of 5) indicates presence of information bias. However this "round digit" or "preferred digit" CPD frequency could also be true reporting. The highest peak at '10' in the reported CPD frequency could be because of availability of manufactured cigarettes in the form of '10 cigarettes' per pack in the market and consumption of one packet per day. A similar explanation was earlier provided by the Klesges et al. (1995) for frequent reporting of 20 as the preferred CPD consumption pattern in USA. While selling of loose cigarettes could promote higher consumption (Latkin et al., 2013), this form of marketing (ten cigarettes per pack) in India could compel smokers to consume all available cigarettes with him in a day. So this point directs towards a policy implication to reduce numbers of cigarettes per pack that may help to reduce

consumption among the smokers. However this hypothesis needs to be explored further.

Subjective self-assessment through global questions depends on participants' abilities to remember and recall their own behavior for the course of 30 days which has been shown to be inaccurate owing to cognitive biases (Hammersley, 1994; Harris et al., 2009; Shiffman, 2009) which may lead to under reporting of CPD frequency (Warner et al., 1978; Pechacek et al., 1984). In view of lack of 'Gold Standard', against which self-reports of CPD frequency could be evaluated (Shiffman, 2009), cross sectional survey need to be more careful about digit bias considering socio-cultural and socio-demographic factors. Saliva cotinine verified study by Dhavan et al. (2011) indicates that self-reports of tobacco use has a low sensitivity (36.3%) in Indian youths of 10-19 years old. ROC analysis of urinary cotinine levels in detection of self reported smoking by Balhara et al (2013) among adult psychiatric patients in North India, yielded area under curve (AUC) of 0.44, while true are is 0.5. Also being a socially undesirable behaviour, tobacco use is prone to under reporting (Fendrich et al., 2005). Hence the possibility of underreporting or misreporting in Indian context is high. Despite of higher possibility of under reporting in females in Indian context, Giovino et al. (2012) reported that mean CPD use among female Indian is higher than their male counterparts. It may be due to misreporting (Jena et al., 2012). The prevalence of higher digit bias among males than females and thus resultant more rounding down of CPD frequency among males may be an explanation for such unusual findings. Further exploration into definition and statistical analysis in GTAs is warranted.

In this study, 2, 5 and 10 have been most frequently reported CPD frequency. Therefore, special attention is required while registering such frequencies in the electronic data base generated though hand held device in the GATS. If necessary, further probing questionnaire may be asked. Since CPD data represents consumption pattern in the population, it could be verified from industrial (IARC, 2008) data, manufacturing data and sales data from government tax department. It is a crude method, but may form an alternative. Carlo et al. (2008) has modeled general pattern of digit preference and has tested it for optimal smoothing of various digit biased data (age at death, weight data etc.). Similar approach may also be used while reporting CPD frequency in large cross sectional surveys like GATS. The analysis of CPD frequency as a categorical data may reduce digit bias but under reporting and rounding down may miss classify the cigarette use. Examination of other methods of data collection which could ensure the accuracy of CPD data is warranted.

In conclusion, present study highlights poor data quality and high prevalence of digit bias in self reported CPD use frequency in India. Reduction of digit bias by innovative methods and statistical analysis should be given due importance in smoking behavioural surveys. The present study needs to be further expanded to investigate the nature of digit bias and the possibility of reduction of consumption of cigarettes among daily smokers by

reducing cigarettes per pack available in the market

Acknowledgements

We are thankful to Centre for Disease Control, Atlanta for freely availing making the GATS-India data in the public domain.

References

- Balhara YP, Jain R (2013). A receiver operated curve-based evaluation of change in sensitivity and specificity of cotinine urinalysis for detecting active tobacco use. *J Cancer Res Ther*, **9**, 84-9.
- Berkman ET, Dickenson J, Falk EB, et al (2011). Using SMS text messaging to assess moderators of smoking reduction: validating a new tool for ecological measurement of health behaviors. *Health Psychol*, **30**, 186-94.
- Caraballo RS, Giovino GA, Pechacek TF, et al (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years or older: Third National Health and Nutrition Examination Survey, 1988-1994. *Am J Epidemiol*, **153**, 807-14.
- Caraballo RS, Giovino GA, Pechacek TF (2004). Self-reported cigarette smoking vs. serum cotinine among US adolescents. *Nicotine Tob Res*, **6**, 19-25.
- Carlo GC, Paul HCE, Jutta G (2008). Modelling general patterns of digit preference. *Statistical Modelling*, **8**, 385-401
- Centers for Disease Control and Prevention (CDC). Global Tobacco Surveillance System data (GTSSData). <http://apps.nccd.cdc.gov/gtssdata/Ancillary/DataReports.aspx?CAID=2> (accessed on Nov 20, 2012).
- Chaiton MO, Cohen JE, McDonald WP, et al (2007). The Heaviness Of Smoking Index as a predictor of smoking cessation in Canada. *Addict Behav*, **32**, 1031-42.
- Crocketta GRM, Crocketts AC, Turner SJ (2001). 'Base-number correlation': a new technique for investigating digit preference and data heaping. *Int J Humanities and Arts Computing*, **13**, 161-79.
- Dawe S, Loxton NJ, Hides L, Kavanagh DJ, Mattick RP (2002). Review of diagnostic screening instruments for alcohol and other drug use and other psychiatric disorders (2nd Ed). Australia: Commonwealth Department of Health and Ageing.
- Denic S, Khatib F, Saadi H (2004). Quality of age data in patients from developing countries. *J Public Health (Oxf)*, **26**, 168-71.
- Dhavan P, Bassi S, Stigler MH, et al (2011). Using salivary cotinine to validate self-reports of tobacco use by Indian youth living in low-income neighborhoods. *Asian Pac J Cancer Prev*, **12**, 2551-4.
- Directorate General of Health Services (DGHS) (2011). Tobacco Dependence Treatment Guidelines. New Delhi: DGHS, Ministry of Health or Family Welfare, Government of India.
- Fagerstrom K (2003). Time to fi rst cigarette; the best single indicator of tobacco dependence? *Monaldi Arch Chest Dis*, **59**, 91-4
- Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbel A, Wislar JS (2005). Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav*, **30**, 175-81.
- Giovino GA, Mirza SA, Samet JM et al (2012). Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. *Lancet*, **380**, 668-79.
- Global Adult Tobacco Survey Collaborative Group (2011). Tobacco Questions for Surveys: A Subset of Key questions

- from the Global Adult Tobacco Survey-GATS (2nd Ed). Atlanta, GA: Centers for Disease Control and Prevention.
- Hammersley R (1994). A digest of memory phenomena for addiction research. *Addiction*, **89**, 283-93.
- Harris KJ, Golbeck AL, Cronk NJ, et al (2009). Timeline follow-back versus global self-reports of tobacco smoking: A comparison of findings with non-daily smokers. *Psychol Addict Behav*, **23**, 368-72.
- Hatziandreu EJ, Pierce JP, Fiore MC, et al (1989). The reliability of self-reported cigarette consumption in the United States. *Am J Public Hlth*, **79**, 1020-23.
- Heatherton TF, Kozlowski LT, Frecker RC, et al (1991). The Fagerstrom test for nicotine dependence: A revision of the Fagerstrom tolerance questionnaire. *Br J Addict*, **86**, 1119-27.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2004). Tobacco smoke and involuntary smoking (IARC Monograph 83). Lyon, France: IARC, World Health Organization.
- IARC Handbooks of Cancer Prevention (2008). Tobacco Control Vol. 12: Methods for Evaluating Tobacco Control Policies. Lyon, France: IARC, World Health Organization.
- International Institute for Population Studies (IIPS) (2010). Global Adult Tobacco Survey (GATS)-India; 2009-10. New Delhi: Ministry of Health or Family Welfare, Government of India.
- Jena PK, Kishore J, Bandyopadhyay C (2012). Prevalence and patterns of tobacco use in Asia. *Lancet*, **380**, 1906.
- Klesges RC, Debon M, Ray JW (1995). Are self-reports of smoking rate biased? Evidence from the second national health and nutrition examination survey. *J Clin Epidemiol*, **48**, 1225-33.
- Latkin CA, Murray LI, Clegg Smith K, et al (2013). The prevalence and correlates of single cigarette selling among urban disadvantaged drug users in Baltimore, Maryland. *Drug Alcohol Depend*, **10**, 007.
- Law MR, Morris JK, Watt HC, Wald NJ (1997). The dose-response relationship between cigarette consumption, biochemical markers and risk of lung cancer. *Br J Cancer*, **75**, 1690-93.
- Means B, Habina K, Swan GE, et al (1992). Cognitive research on response error in survey questions on smoking. Maryland, USA: National Center for Health Statistics, United States Department of Health and Human Services.
- Pardeshi GS (2010). Age heaping and accuracy of age data collected during a community survey in the Yavatmal District, Maharashtra. *Indian J Community Med*, **35**, 391-5.
- Pechucek TF, Fox BH, Murray DM et al (1984). Review of techniques for measurement of smoking behavior. In: Matarazzo JD, Weiss SM, Herd JA, Miller NE, eds. Behavioral health: A handbook of health enhancement and disease prevention. New York: Wiley.
- Perkins KA, Jao NC, Karelitz JL (2012). Consistency of daily cigarette smoking amount in dependent adults. *Psychol Addict Behav*, **10**, 30287.
- Shiffman S (2009). Commentary on Herd or Borland (2009) and Herd et al. (2009): Illuminating the course and dynamics of smoking cessation. *Addiction*, **104**, 2100-01.
- Shiffman S (2009). How many cigarettes did you smoke? Assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment. *Hlth Psychol*, **28**, 519-26.
- US Department of Health and Human Services (2008). Clinical Practice Guideline - Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: US Dept of Health and Human Services, Public Health Service.
- Wang Z, Zeng Y, Jeune B, et al (1998). Age validation of Han Chinese centenarians. *Genus*, **54**, 123-41.
- Warner KE (1978). Possible increases in the underreporting of cigarette consumption. *J Am Stat Assn*, **73**, 314-18.
- West R, McNeill A, Raw M (2000). Smoking cessation guidelines for health professionals: an update. *Thorax*, **55**, 987-99.