

Association between socioeconomic status and self-reported diabetes in India: a cross-sectional multilevel analysis

Daniel J Corsi,¹ S V Subramanian²

To cite: Corsi DJ, Subramanian SV. Association between socioeconomic status and self-reported diabetes in India: a cross-sectional multilevel analysis. *BMJ Open* 2012;**2**: e000895. doi:10.1136/bmjopen-2012-000895

► Prepublication history and additional tables for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-000895>).

Received 18 January 2012
Accepted 18 June 2012

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

¹Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada

²Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, Massachusetts, USA

Correspondence to

Professor S V Subramanian; svsubram@hsph.harvard.edu

ABSTRACT

Objectives: To quantify the association between socioeconomic status (SES) and type 2 diabetes in India.

Design: Nationally representative cross-sectional household survey.

Setting: Urban and rural areas across 29 states in India.

Participants: 168 135 survey respondents aged 18–49 years (women) and 18–54 years (men).

Primary outcome measure: Self-reported diabetes status.

Results: Markers of SES were social caste, household wealth and education. The overall prevalence of self-reported diabetes was 1.5%; this increased to 1.9% and 2.5% for those with the highest levels of education and household wealth, respectively. In multilevel logistic regression models (adjusted for age, gender, religion, marital status and place of residence), education (OR 1.87 for higher education vs no education) and household wealth (OR 4.04 for richest quintile vs poorest) were positively related to self-reported diabetes ($p < 0.0001$). In a fully adjusted model including all socioeconomic variables and body mass index, household wealth emerged as positive and statistically significant with an OR for self-reported diabetes of 2.58 (95% credible interval (CrI): 1.99 to 3.40) for the richest quintile of household wealth versus the poorest. Nationally in India, a one-quintile increase in household wealth was associated with an OR of 1.31 (95% CrI 1.20 to 1.42) for self-reported diabetes. This association was consistent across states with the relationship found to be positive in 97% of states (28 of 29) and statistically significant in 69% (20 of 29 states).

Conclusions: The authors found that the highest SES groups in India appear to be at greatest risk for type 2 diabetes. This raises important policy implications for addressing the disease burdens among the poor versus those among the non-poor in the context of India, where >40% of the population is living in poverty.

ARTICLE SUMMARY

Article focus

- The relationship between socioeconomic factors and type 2 diabetes has not been previously studied for the whole of India and across states.
- Our objective was to investigate associations between measures of SES (defined as social caste, education, household wealth) and self-reported diabetes status in India.
- In addition, we explored geographic variation in the prevalence of diabetes between states and local areas in India and between-state variability in the SES–diabetes relationship.

Key messages

- The highest socioeconomic groups appear to be at greatest risk for diabetes in India with the strength of the association consistent in size and magnitude across states.
- There is substantial geographic heterogeneity in the prevalence of diabetes.
- These findings raise important policy implications for addressing the disease burdens among the poor versus those among the non-poor in the context of India, where nearly half of the population is living in poverty.

Strengths and limitations of this study

- The key strength of this study is the use of a large nationally representative survey to assess the socioeconomic and geographic patterning of diabetes across all of India. Limitations include the relatively younger age of the sample and assessment of diabetes status on the basis of self-reports.

INTRODUCTION

The prevalence of type 2 diabetes in India has been investigated in numerous population-based surveys conducted across a range of settings since the 1970s.^{1–6} Despite multiple prevalence studies, no nationally representative studies exist that have considered the association between socioeconomic status (SES) and type 2 diabetes in India. In

a review of 15 existing studies that have reported the prevalence of type 2 diabetes by SES and/or associations between SES and type 2 diabetes, all were found to have been based on local or regional samples and a majority were done in urban areas^{4 6–19} (table 1). It has been suggested that the prevalence of type 2 diabetes and other cardiovascular disease risk factors may increasingly become concentrated among low SES groups in India²⁰ and other low- and middle-income countries,²¹ although to date the empirical evidence from India in support of this hypothesis remains limited. The majority of studies reviewed in table 1 have provided evidence of a positive association between SES (defined as education, household wealth, social caste or a composite of two or more markers) and diabetes among populations from selected geographic regions in India^{6 11 17}; however, the strength and consistency of this association across the whole of India has not previously been assessed.

Type 2 diabetes is the most common form of diabetes globally, accounting for >85% of cases.²² The incidence of type 2 diabetes is related to genetic and non-genetic components, with the latter being greatly influenced by modifiable risk factors such as obesity, diets low in fibre and high in trans fat and physical inactivity.^{23 24} Lifestyle behaviours are strongly patterned by SES²⁵ and may be mediators on the causal pathway between SES and the onset of type 2 diabetes.²⁶ In high-income countries, the SES–diabetes relationship appears to be negative, with the poor at greatest risk. For example, strong associations have been observed between poverty, low education and type 2 diabetes among African–American women^{27 28} and among White women and men in the USA.²⁹ Similarly, a study from Canada described an inversely graded SES–diabetes association with an OR of 1.9 for men (95% confidence interval (CI) 1.6 to 2.4) and 2.8 (95% CI 2.2 to 3.4) for women for the lowest versus highest income groups.³⁰ A recent meta-analysis of 23 case–control and cohort studies and 43 measures of SES–diabetes association revealed an overall increased risk for type 2 diabetes for low SES groups based on education, occupation and income.³¹ The strength of the association, however, was less consistent in low- and middle-income countries, and few studies have been conducted in these countries.

Concern has been raised over the anticipated rapid increase in type 2 diabetes prevalence in India.^{32 33} Evidence on the secular increases in diabetes prevalence in India, however, have been limited to urban areas of Southern India^{4 34 35} and have focused on the mean rates of diabetes rather than how it is distributed in the population. In this paper, we address the need to comprehensively investigate the socioeconomic and geographic distribution of type 2 diabetes in the Indian population using a large-scale nationally representative survey. Specifically, we investigate the SES–diabetes association through the SES markers of social caste, household wealth and education. In addition, we investigate the geographic distribution of the prevalence of

diabetes across states and local areas along with variability in the SES–diabetes association across states.

METHODS

Data source

We use data from the 3rd National Family Health Survey (NFHS), conducted in 29 states in India between November 2005 and August 2006.³⁶ NFHS-3 is a major national health survey in India that collected information on a range of indicators including reproductive health, nutritional status of adults and children, utilisation of healthcare services and blood testing for HIV prevalence. NFHS-3 covered all states in India, which comprises nearly 99% of the population, but excluded Union Territories. The survey was designed to provide estimates of key indicators (except HIV prevalence) for each state by urban and rural areas.

Survey design

A uniform multistage sampling strategy was adopted in all states, with separate sampling in urban and rural areas.^{36 37} In rural areas, a two-stage sample was carried out using a list of villages from the 2001 census as the sampling frame. In the first stage, a stratified sample of villages was drawn with probability proportional to the size of the village. In the second stage, a random selection of households was drawn in each village from a complete list of households compiled during field visits carried out in each sampled village. In urban areas, a similar procedure was implemented beginning with a stratified random sample of municipal wards based on the 2001 census. Next, one census enumeration block (150–200 households) was selected from within wards using probability proportional to size. Finally, as in rural areas, field enumerators undertook a household listing operation in selected blocks and a random sample of households was made. In both rural and urban areas, 30 households were targeted for selection in each of the sampled units. The overall household response rate for NFHS-3 was 98%.³⁶

All women aged 15–49 years in selected households were invited to participate in the survey. In 22 states, men aged 15–54 years in a random subsample of households drawn from each PSU (about six households per PSU) were eligible for the men's survey. In the remaining seven states (Andhra Pradesh, Karnataka, Maharashtra, Manipur, Tamil Nadu, Uttar Pradesh and Nagaland), eligible men in all selected households were invited to participate. The additional men recruited in these states was for the purpose of HIV testing to provide reliable state-level estimates of HIV prevalence in certain states. Interviews were conducted in one of the 18 Indian languages in the respondent's home and the response rates were 95% for women and 87% for men.³⁶ During interviews, the weights and heights of survey respondents were measured by trained field technicians using standardised measuring equipment designed for survey settings.³⁸

Table 1 Overview of studies reporting prevalence of type 2 diabetes by markers of SES and the association between increasing SES and diabetes in India

First author	Study period	Coverage	Setting	Age	Sample size	Diabetes assessment	SES marker	Gender	Diabetes prevalence:		SES–diabetes association: OR (95% confidence interval) for high SES versus low SES
									low SES (l); high SES (h)	low SES	
Singh ⁷	1994	Local	Rural	25–64	1769	Blood glucose	Composite	Male	0.9% (l); 6.1% (h)*	–	–
Singh ⁸	1994	Local	Rural	25–64	1806	Blood glucose	Composite	Female	0.9% (l); 6.9% (h)*	–	–
								Male	2.5% (l); 8.6% (h)*	2.03 (1.86 to 2.51)*	–
Singh ⁹	1994	Local	Combined	25–64	3575	Blood glucose	Composite	Female	1.2% (l); 6.9% (h)*	–	1.97 (1.67 to 2.36)*
								Male	–	4.07 (1.89 to 10.01)* (Urban)	–
Singh ¹⁰	1998	Regional	Urban	25–64	3257	Blood glucose	Composite	Female	0.5% (l); 4.8% (h)*	–	–
								Combined	12.5% (l); 21.6% (h)*	1.43 (1.30 to 1.57)*;	–
Ramachandran ⁴	2000	Regional	Urban	20+	11 216	Blood glucose	Income	Combined	12.6% (l); 25.5% (h)*	–	2.15 (1.70 to 2.72)
Ramachandran ¹¹	1999–2000	Local	Urban	40+	2383	Blood glucose, drug treatment	Income	Combined	–	–	–
Gupta ¹²	1999–2001	Local	Urban	20+	1123	Self-report	Education	Male	6.8% (l); 7.9% (h)	–	–
								Female	6.6% (l); 8.3% (h)	–	–
Reddy ¹³	2002–2003	Regional	Urban	20–69	19 973	Blood glucose, drug treatment	Education	Male	7.6% (l); 8.4% (h)	–	1.11 (0.71 to 1.67)
								Female	–	–	–
Mohan ⁶	2003–2005	Regional	Combined	15–64	44 523	Self-report	Education	Female	11.2% (l); 4.2% (h)*	–	0.36 (0.23 to 0.56)*
								Combined	3.4% (l); 5.6% (h)*	–	3.02 (2.45 to 3.71)*
Ajay ¹⁴	2002–2003	Regional	Urban	20–69	10 930	Blood glucose, drug treatment	Education	Combined	11.6% (l); 6.9% (h)*	–	0.69 (0.54 to 0.89)*
								Female	–	–	–
Vijayakumar ¹⁵	2007	Local	Rural	18+	1990	Blood glucose, self-report	Social caste	Combined	5.9% (l); 17.4% (h)	–	–
Gupta ¹⁶	1999–2003	Local	Urban	20–59	1289	Blood glucose, self-report	Wealth	Combined	8.0% (l); 18.8% (h)*	–	1.43 (1.04 to 1.95)*
								Male	–	–	–
Kinra ¹⁷	2005–2007	Regional	Rural	20–69	1983	Blood glucose, self-report	Wealth	Female	6.0% (l); 34.7% (h)*	–	–
								Combined	6.9% (l); 26.4% (h)*	–	–
								Male	1.8% (l); 8.0% (h)*	–	–
								Female	3.9% (l); 5.1% (h)	–	–

Continued

Table 1 Continued

First author	Study period	Coverage	Setting	Age	Sample size	Diabetes assessment	SES marker	Gender	Diabetes prevalence: low SES (l); high SES (h)	SES—diabetes association: OR (95% confidence interval) for high SES versus low SES
Samuel ¹⁸	1969–2002	Regional	Urban	26–32	2218	Blood glucose,† drug treatment	Wealth	Male	26.2% (l); 31.9% (h)*	—
			Rural					Female	12.1% (l); 30.3% (h)*	—
			Combined					Male	10.9% (l); 31.8% (h)*	—
			Urban					Female	16.1% (l); 32.1% (h)*	—
			Rural					Combined	—	2.8 (1.9 to 4.1)*
Zaman ¹⁹	2005	Regional	Urban	30+	4535	Blood glucose, self-report	Income	Male	15.0% (l); 34.7% (h)*	—
			Rural					Female	31.5% (l); 32.2% (h)	—
			Combined					Male	25.7% (l); 19.7% (h)	—
			Rural					Female	19.1% (l); 50.0% (h)	—
			Combined					Combined	—	1.0 (0.6 to 1.6)

SES markers defined as education, household wealth, social caste or a composite of two or more measures.

*p<0.05.

—, Indicates not reported; SES, socioeconomic status.

† Includes impaired glucose tolerance and impaired fasting glucose.

In total, NFHS-3 collected information from 109 041 households, 124 385 women aged 15–49 years and 74 369 men aged 15–54 years. We restricted our analyses to adults aged ≥ 18 years and non-pregnant women ($n=171\,207$). Respondents who did not report or know their diabetes status ($n=2373$) or with incomplete information for any of the independent variables considered in the analysis (marital status, religion, caste, education, household wealth) were excluded ($n=699$). Main analyses were conducted on a sample of 168 135 respondents (65 255 men and 102 880 women). Additional analyses considering body mass index (BMI) were restricted to a sample of 158 936 due to missing and/or implausible values for height and/or weight. Figure 1 provides a flow diagram detailing the NFHS sample, exclusions and final analytic sample sizes.

Outcome and independent variables

The primary outcome was diabetes, assessed on the basis of self-reports by survey respondents. Markers of SES were social caste, household wealth and education. Social caste was reported by the household head. The categories were other caste, scheduled caste, scheduled tribe, other backward class and no caste. Other caste is a heterogeneous group that is traditionally viewed as having higher social status. Scheduled castes and scheduled tribes are considered lower, socially marginalised groups in India.³⁹ Household wealth was defined by an index based on indicators of asset ownership and housing characteristics.⁴⁰ This index has been developed and validated in a number of countries to be a robust measure of wealth and has been found to be consistent with measures of income and expenditure.⁴¹ Briefly, the measure was constructed as follows. Information on 33

indicators of housing characteristics (eg, type of windows and flooring, water and sanitation facilities) and assets (eg, ownership of home, car, computer, mobile phone) was weighted and combined with weights derived from a principal component analysis procedure.³⁶ The resulting variable was standardised to a mean of 0 and SD of 1, and using this index, the household population was divided into fifths from poorest to richest. Education was categorised into four levels as no education, primary, secondary or higher education.

Background characteristics included age, gender, religion, marital status, place of residence and BMI. Age was defined in 10-year categories and centred about its mean (32 years) in regression models. Gender was based on self-report. Religion was categorised as Hindu, Muslim, Sikh, Buddhist or other religion. Marital status was defined as single, married, widowed or divorced/separated. Place of residence (rural or urban) was defined according to the 2001 Census. BMI (in kg/m^2 ; weight in kilogrammes divided by the square of height in metres) was calculated for all survey respondents with valid measurements for weight and height. BMI was classified according to the following categories based on risk of type 2 diabetes and cardiovascular disease in Asian populations: $<18.5 \text{ kg}/\text{m}^2$ (underweight), $18.5\text{--}22.9 \text{ kg}/\text{m}^2$ (acceptable risk), $23\text{--}27.4 \text{ kg}/\text{m}^2$ (increased risk) and $\geq 27.5 \text{ kg}/\text{m}^2$ (high risk).⁴²

Analysis

To account for the complex survey design, we employed multilevel logistic regression to model the probability of diabetes.⁴³ A three-level model was specified with a binary response (y , diabetes or not) for individual i in local area j (village or census block primary sampling units), in state k . The outcome diabetes, $\text{Pr}(y_{ijk}=1)$, was assumed to be binomially distributed $y_{ijk} \sim \text{Binomial}(1, p_{ijk})$ with probability π_{ijk} related to the set of independent variables X and a random effect for each level by a logit link function:

$$\text{Logit}(\pi_{ijk}) = \beta_0 + \beta X_{ijk} + (v_{0k} + u_{0jk}) \quad (\text{Equation 1})$$

The right-hand side of the equation consists of the fixed part linear predictor ($\beta_0 + \beta X_{ijk}$) and random intercepts attributable to states (v_{0k}) and local areas (u_{0jk}). The intercept, β_0 , represents the log odds of diabetes in the reference group, and the β -coefficients represent the differential in the log odds of diabetes compared with the reference group defined for each independent variable. Coefficients were exponentiated and presented as ORs for interpretation. The random intercepts are assumed to be independently and identically distributed and have variances estimated for states (σ_v^2) and local areas (σ_u^2).⁴⁴ The variance parameters quantify heterogeneity in the log odds of diabetes at each level, after taking into account individual characteristics and place of residence in the fixed part. We expressed the variances at each level as a percentage of their contribution to the total variance from an initial

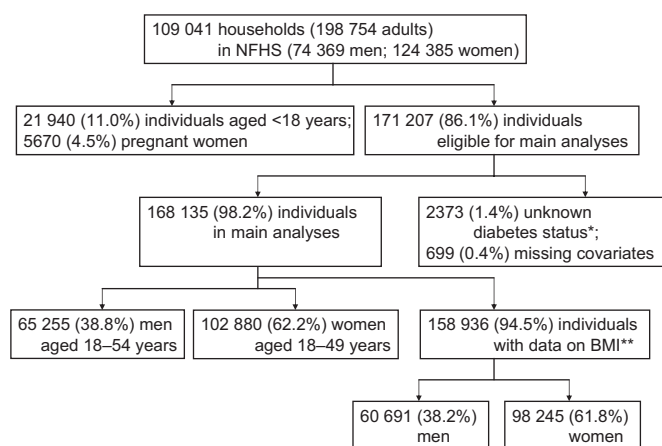


Figure 1 Flow diagram showing exclusions and final sample sizes, 2005–2006 National Family Health Survey (NFHS).

*Two thousand three hundred and thirty-three individuals reported unknown diabetes status; in 40 individuals, diabetes status was not reported/missing. Of the 2333 individuals who reported unknown diabetes status, 2210 (94.7%) had complete data for BMI and were included in sensitivity analyses.

**Analyses involving body mass index (BMI) as an independent variable were restricted to 158 936 individuals.

model adjusting for age and gender only and from a final model accounting for all covariates. We specified a sequence of six models during analyses. In the first three models, one SES marker (social caste, household wealth and education) was added to a model that adjusted for background characteristics (age, gender, religion and place of residence). In the fourth mutually adjusted model, all SES markers were included along with background characteristics from the previous models. In the fifth model, BMI was included with markers of SES and background characteristics from model 4. In the sixth model, we also tested whether the association between household wealth varied across states in terms of strength or direction, given that different states vary tremendously by levels of economic development and could be considered at different levels of epidemiological transition. In order to test this between-state variability, we expanded equation 1 to allow the slope of household wealth to vary across states:

$$\text{Logit}(\pi_{ijk}) = \beta_0 + \beta_{1k} \text{wealth}_{ijk} + \beta X_{ijk} + (v_{0k} + v_{1k} + u_{0jk})$$

(Equation 2)

The key feature of equation 2 is that the effect of wealth on self-reported diabetes in state *k* consists of the overall average effect across all states (β_1), plus a state-specific (v_{1k}) differential in this effect. We summarised and presented the results of this model as the OR for self-reported diabetes overall in India and for each state given a 1-quintile increase in household wealth and conditional on all covariates from model 5. Additional analyses were carried out separately for male and female samples using an identical sequence of models (with the exclusion of gender as a background characteristic). Estimation of models was done using Markov Chain Monte Carlo simulation and the statistical software *MLwiN*.^{45 46}

RESULTS

Characteristics of survey respondents by their self-reported diabetes status are given in table 2. The overall prevalence of diabetes in this sample was 1.5%, and this was higher in urban areas and among men (diabetes prevalence 2.0% in urban vs 1.0% in rural; 1.8% in men vs 1.3% in women). Diabetes prevalence increased with age (7.5% in 50–54 years vs 0.3% in 18–29 years), education (1.9% in higher vs 1.0% in no education), household wealth (2.5% in richest vs 0.4% in poorest) and BMI (4.8% in $\geq 27.5 \text{ kg/m}^2$ vs 0.6% in $< 18.5 \text{ kg/m}^2$). At the state level, the prevalence of diabetes varied between 0.3% in Rajasthan and 3.3% in Kerala and was generally higher in Southern and Eastern states (figure 2).

In separate models that adjusted for age, marital status, religion and place of residence, statistically significant associations were observed between SES and

Table 2 Characteristics of survey participants and frequency distribution of self-reported diabetes in India, males and females from the 3rd National Family Health Survey

	Self-reported diabetes, n (%)	Total, n
Participants	2439 (1.5)	168 135
Residence		
Rural	818 (1.0)	86 013
Urban	1621 (2.0)	82 122
Age group (years)		
18–29	266 (0.3)	76 174
30–39	602 (1.2)	51 132
40–49	1238 (3.4)	36 402
50–54	333 (7.5)	4427
Gender		
Male	1144 (1.8)	65 255
Female	1295 (1.3)	102 880
Marital status		
Single	132 (0.3)	38 078
Married	2165 (1.8)	123 457
Widowed	108 (2.5)	4320
Divorced or separated	34 (1.5)	2280
Religion		
Hindu	1775 (1.4)	123 411
Muslim	340 (1.6)	21 510
Christian	213 (1.4)	14 779
Sikh	49 (1.5)	3236
Buddhist	34 (1.4)	2451
Other	28 (1.0)	2748
Social caste		
Other caste	1026 (1.8)	56 063
Scheduled caste	349 (1.3)	27 677
Scheduled tribe	167 (0.8)	21 372
Other backward class	781 (1.4)	55 641
No caste	116 (1.6)	7382
Education		
No education	464 (1.0)	44 856
Primary	358 (1.4)	24 969
Secondary	1166 (1.6)	74 715
Higher	451 (1.9)	23 595
Household wealth		
Poorest	77 (0.4)	17 252
2nd quintile	175 (0.8)	22 948
3rd quintile	278 (0.9)	32 070
4th quintile	573 (1.4)	42 091
Richest	1336 (2.5)	53 774
Body mass index (kg/m ²)		
<18.5	243 (0.6)	42 128
18.5–22.9	703 (0.9)	74 089
23–27.4	833 (2.7)	31 217
≥ 27.5	547 (4.8)	11 502

self-reported diabetes for each of the primary markers of SES in this study: social caste, household wealth and education. Compared with the other caste group, scheduled casts, scheduled tribes and other backward classes had reduced odds of having diabetes with ORs of 0.81 (95% CrI 0.71 to 0.94), 0.57 (95% CrI 0.46 to 0.70) and 0.84 (95% CrI 0.75 to 0.94), respectively (table 3,

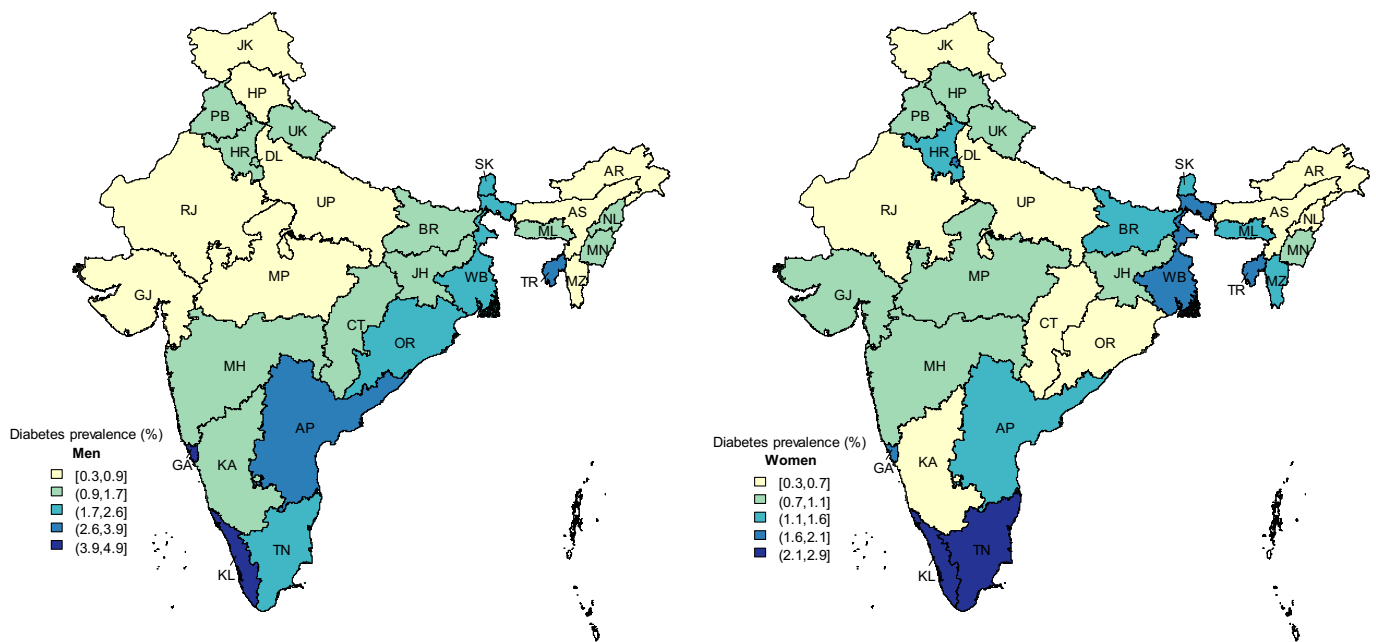


Figure 2 State-level prevalence of self-reported diabetes in India for men aged 18–54 years (left) and women aged 18–49 years (right). Darker colours indicate higher prevalence. State name abbreviations: AP, Andhra Pradesh; AR, Arunachal Pradesh; AS, Assam; BR, Bihar; CT, Chhattisgarh; DL, Delhi; GA, Goa; GJ, Gujarat; HR, Haryana; HP, Himachal Pradesh; JK, Jammu & Kashmir; JH, Jharkhand; KA, Karnataka; KL, Kerala; MP, Madhya Pradesh; MH, Maharashtra; MN, Manipur; ML, Meghalaya; MZ, Mizoram; NL, Nagaland; OR, Orissa; PB, Punjab; RJ, Rajasthan; SK, Sikkim; TN, Tamil Nadu; TR, Tripura; UP, Uttar Pradesh; UK, Uttarakhand (Uttaranchal); WB, West Bengal.

models 1–3). Education showed a graded relationship with diabetes and an OR of 1.87 (95% CrI 1.61 to 2.18) for those with higher education versus those with no education. Household wealth showed a graded association with diabetes with individuals from the richest households having an OR for diabetes of 4.04 (95% CrI 3.08 to 5.30) compared with those from the poorest households.

The effects of social caste and education were attenuated in the mutually adjusted model (model 4), suggesting that their independent effects on self-reported diabetes were at least partially mediated by the inclusion of household in this model. The reduced odds for diabetes remained consistent for scheduled tribes versus other caste groups (OR 0.72, 95% CrI 0.58 to 0.90) as did an increased odds for those with secondary education versus no education (OR 1.18, 95% CrI 1.04 to 1.35); however, the graded relation with education was less consistent. In separate mutually adjusted models that were stratified by gender, education showed a graded association in men although it was not statistically significant with the OR for diabetes men found to be 1.27 (95% CrI 0.98 to 1.70) for men with higher versus no education (supplemental table 1). Among women, those with secondary education continued to show an increased odds of self-reported diabetes compared with those with no education (OR 1.28, 95% CrI 1.08 to 1.50). Overall, the strong and graded relationship between household wealth and diabetes remained consistent in model 4 with an OR for diabetes of 3.65 (95% CrI 2.83 to

4.78) for the richest versus the poorest groups; similar associations were found in the gender-specific models (supplemental table 1). Type 2 diabetes is strongly influenced by body weight.^{47–49} Therefore, BMI was added to model 5 to control for potential confounding of the SES–diabetes relationship in this sample. In addition, BMI was added separately in this model because its inclusion resulted in the reduction of sample size by ~5% due to missing values for BMI. The ORs for caste and education remained consistent between the mutually adjusted model and final model that included BMI. The ORs for household wealth were further attenuated in the final model; however, the positive graded association remained statistically significant with an adjusted OR for those in the richest compared with the poorest households of 2.58 (95% CrI 1.99 to 3.40).

Our analyses revealed dramatic variation in the prevalence of diabetes between states and local areas in India (table 4). In an initial multilevel model adjusted for age and gender, states and local areas (defined as villages in rural areas and census blocks in urban areas) contributed 5.9% and 10.8%, respectively, to the total variation in diabetes. The addition of socioeconomic and demographic characteristics along with BMI to the model reduced, the variance in diabetes attributed to local areas by 41%–6.4% but the variation attributed to states was relatively unchanged at 5.4%.

Overall in India, the log odds for diabetes for the reference category (a 32-year-old married women, with no education, BMI <18.5 kg/m², belonging to the other

Table 3 Associations between socioeconomic status and self-reported diabetes in India; 3rd National family health survey, 2005–2006

Variable	Models 1–3, OR (95% CrI)	Model 4, OR (95% CrI)	Model 5, OR (95% CrI)
Social caste			
Other caste	1.00	1.00	1.00
Scheduled caste	0.81 (0.71 to 0.94)	1.05 (0.91 to 1.21)	1.07 (0.93 to 1.24)
Scheduled tribe	0.57 (0.46 to 0.70)	0.72 (0.58 to 0.90)	0.73 (0.57 to 0.92)
Other backward caste	0.84 (0.75 to 0.94)	0.95 (0.85 to 1.07)	0.96 (0.86 to 1.08)
No caste	0.89 (0.71 to 1.11)	0.94 (0.75 to 1.17)	0.95 (0.76 to 1.20)
Wealth			
Poorest	1.00	1.00	1.00
2nd quintile	1.59 (1.20 to 2.12)	1.57 (1.21 to 2.07)	1.49 (1.14 to 1.96)
3rd quintile	1.63 (1.23 to 2.16)	1.55 (1.21 to 2.02)	1.39 (1.07 to 1.81)
4th quintile	2.42 (1.85 to 3.17)	2.25 (1.76 to 2.92)	1.79 (1.40 to 2.34)
Richest	4.04 (3.08 to 5.30)	3.65 (2.83 to 4.78)	2.58 (1.99 to 3.40)
Education			
No education	1.00	1.00	1.00
Primary	1.23 (1.06 to 1.43)	1.06 (0.91 to 1.22)	1.00 (0.86 to 1.17)
Secondary	1.68 (1.49 to 1.90)	1.18 (1.04 to 1.35)	1.12 (0.98 to 1.28)
Higher	1.87 (1.61 to 2.18)	1.12 (0.95 to 1.32)	1.01 (0.86 to 1.20)
Body mass index (kg/m²)			
<18.5			1.00
18.5–22.9			1.25 (1.08 to 1.46)
23–27.4			2.08 (1.79 to 2.44)
≥27.5			2.98 (2.51 to 3.54)

In models 1–3, one SES marker (social caste, household wealth and education) was modelled at a time while adjusting for age, gender, religion and place of residence. In model 4, all SES markers were included along with covariates from models 1–3. In model 5, BMI was included with markers of SES and covariates from model 4. SES, socioeconomic status.

caste group, in the poorest fifth of households and living in a rural area) was –6.13 or a 0.22% probability of diabetes. Compared with this national reference point, being a resident of several Southern and Northeastern states was associated with a statistically significant increase in the odds of diabetes (figure 3). The ORs for self-reported diabetes of these states were 2.29 (Tripura), 1.69 (Tamil Nadu), 1.69 (Kerala), 1.71 (Goa), 1.49 (Andhra Pradesh) and 1.56 (West Bengal). In contrast being resident of the states of Rajasthan, Jammu & Kashmir, Uttar Pradesh, Punjab, Madhya Pradesh and Assam in Northern and Central India was associated with a statistically significant decrease (OR<1.0) in the odds of self-reported diabetes.

In order to assess the variability in the SES–diabetes association across states in India, a final model (model 6) was specified to allow the OR for diabetes for a one-

quintile increase in household wealth to vary across states. In this model, the overall OR for diabetes in India for a one-quintile increase in household wealth was 1.31 (95% CrI 1.20 to 1.42) (figure 4). In 15 states, the association was stronger than the national average, varying between an OR of 1.33 in Rajasthan and 1.55 in Jammu & Kashmir. Although the association was less than the national average in 14 states, it was found to be positive in 28/29 (97%) states and statistically significant in 20/29 (69%). Only in West Bengal was an inverse association observed, but it was not statistically significant (OR 0.95, 95% CrI 0.83 to 1.09). ORs and 95% CrI for the overall association and across all states are presented in supplemental table 2. In summary, the association between household wealth and self-reported diabetes was consistent across the states both in direction and magnitude.

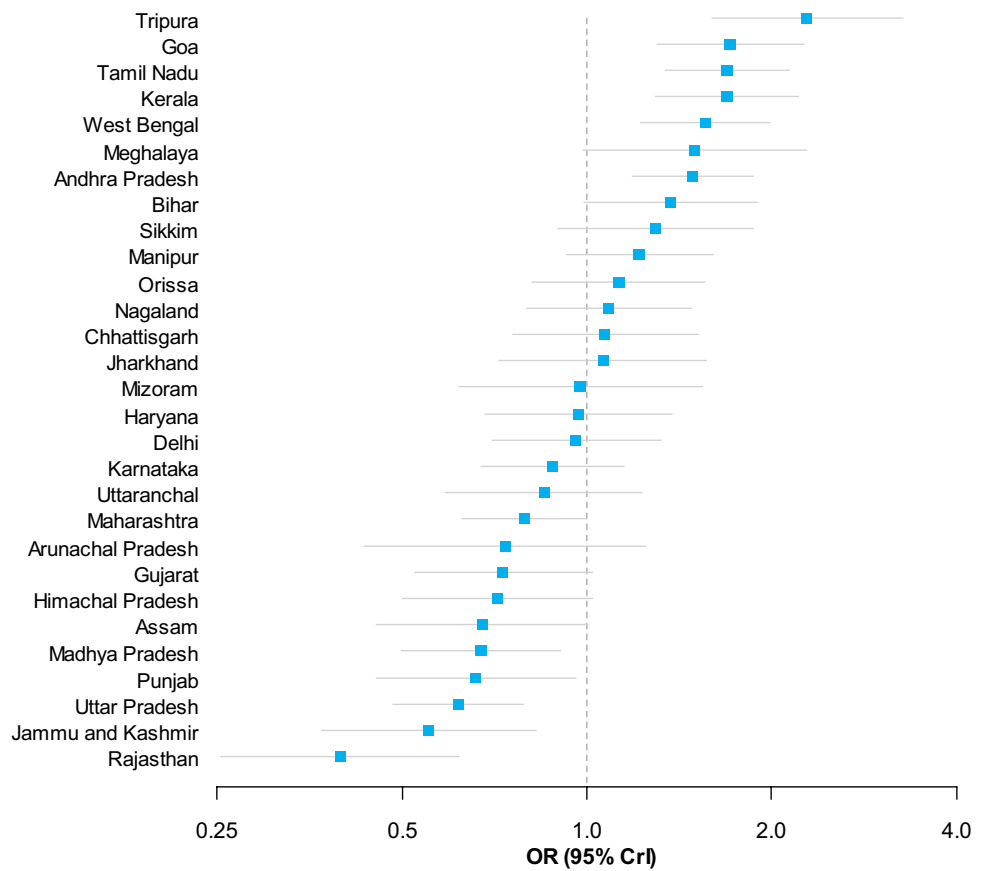
Table 4 Variance in self-reported diabetes status between local areas and states in India, expressed as percentage of the contribution to the total variance in diabetes

	Age and gender adjusted*			Fully adjusted†		
	Variance	SE	%	Variance	SE	%
States	0.231	0.076	5.9	0.204	0.068	5.4
Local areas	0.425	0.043	10.8	0.240	0.041	6.4

*Multilevel model adjusted for age and gender only.

†Multilevel model fully adjusted for age, gender, marital status, religion, social caste, household wealth, education, body mass index and place of residence.

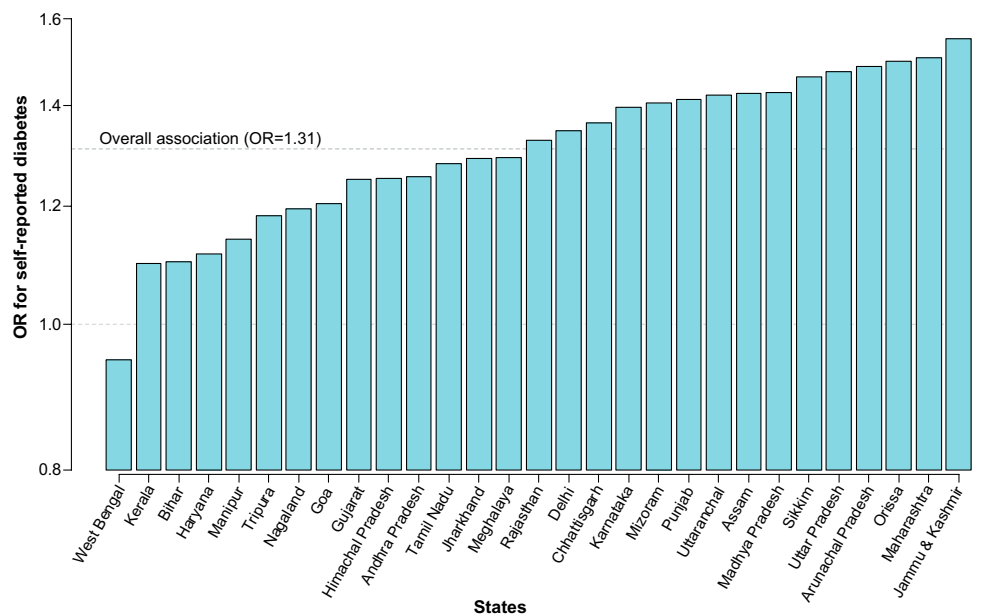
Figure 3 ORs for self-reported diabetes by state of residence in India. Horizontal lines are 95% credible intervals; adjusted for age, gender, marital status, religion, social caste, household wealth, education, body mass index and place of residence.



We conducted several sensitivity analyses to assess the consistency of our findings. First, we examined whether the observed associations were related to respondents' awareness and knowledge about diabetes. To do so, we considered responses to the question, 'Do you have diabetes?' as a categorical variable, comparing 'yes' (diabetic) and 'don't know' (unknown) versus 'no' (non-diabetic) across the same set of independent vari-

ables using a multinomial logistic model. Associations between SES variables and positive reports of diabetes from this model, which included the possibility that respondents were unaware of their diabetes status, were nearly identical to findings from the logistic model that excluded those with unknown diabetes status (supplemental table 3). The multinomial model also revealed that the richer and more highly educated respondents

Figure 4 OR for self-reported diabetes for a one-quintile increase in household wealth for men (aged 18–54 years) and women (aged 18–49 years) in India and 29 states. Adjusted for age, gender, marital status, religion, social caste, education, body mass index and place of residence.



were less likely to report unknown status compared with non-diabetic. In addition, we examined BMI across the three categories of diabetes status (figure 5). This revealed that those with unknown diabetes had the lowest BMI (mean 20.9, SD 3.7) which was largely consistent with the non-diabetic group (mean 21.1, SD 3.9) and substantially lower than those with self-reported diabetes (mean 24.4, SD 4.9). Finally, we examined interactions between socioeconomic variables (caste, education and wealth) and diabetes by residential location. Tests of these interactions were not statistically significant ($p=0.20$ for caste; $p=0.72$ for education; $p=0.66$ for wealth).

DISCUSSION

In this study, we have three key findings. First, measures of SES were positively associated with self-reported diabetes in the NFHS-3. Although the observed effects of caste and education were largely attenuated in fully adjusted models, the effect of household wealth remained positive, graded and statistically significant even after controlling for BMI. Second, we observed a large variation in the prevalence of diabetes between local areas and states in India. A few southern and northeastern states were associated with a higher risk for reporting diabetes while several northern and central states were at lower risk after adjusting for individual characteristics and place of residence. Lastly, the observed association between household wealth

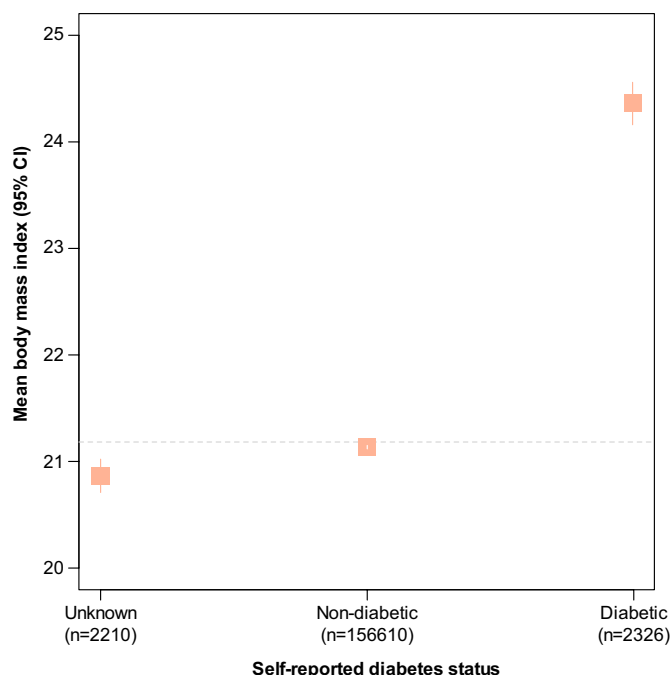


Figure 5 Mean body mass index (BMI) across three possible responses for self-reported diabetes. Vertical lines represent 95% confidence intervals. BMI (in kg/m²) was calculated from measured height and weight values. Horizontal line represents overall mean BMI (21.2 kg/m², SD 3.9).

and self-reported diabetes was consistent, positive, and statistically significant across a majority of states in India.

There are a few limitations to our study. First, the outcome was defined on the basis of self-reported diabetes, although interviews were conducted in person using a standardised instrument. Previous research has shown good agreement for self-reported diabetes when compared with medical records in a US population⁵⁰ and that self-reported health conditions demonstrate the expected relationship with SES in India.⁵¹ In addition, our sensitivity analyses considering respondents who reported 'unknown' for diabetes status were nearly identical to the main analyses. We did find, however, evidence that higher SES groups were less likely to report 'did not know' as compared with 'no', which has been suggested previously on studies using self-reports of diabetes status in India.⁶ However, the unknown group was more similar in terms of BMI, education and wealth to the non-diabetic rather than diabetic group. In addition, our findings of positive SES–diabetes associations were consistent with several studies identified in our literature review that used blood glucose measurements for the assessment of diabetes status (summarised in table 1). Lastly, although our sample was relatively young (<55 years for men and <50 years for women), it is representative of the young population of profile of India; 84% of the Indian adult population (18–69 years) and 47% of the total Indian population at all ages fall within the ages covered by this study.⁵² Our study does exclude approximately 12% of the Indian population (women over the age of 50 and men over the age of 55) due to the sample design of the NFHS. The prevalence of diabetes increases with age and whether a similar SES–diabetes relationship exists among middle and older age groups in all parts India is not clear, although our findings are consistent with the previous studies that have included older ages.

Our findings of positive SES–diabetes associations are consistent with the previous studies done in different parts of India. For example, an analysis of rural participants from the Indian Migration Study, which sampled primarily from four large states in the north, centre and south of India,¹⁷ identified a positive SES–diabetes gradient among men (8.0% prevalence in high SES group vs 1.8% in low SES group) and a weaker positive SES–diabetes association that was not statistically significant among women (5.1% vs 3.9%). In addition, a study done in an urban setting in Madras (Chennai) found an OR for diabetes of 2.2 (95% CI 1.7 to 2.7) for high versus low SES groups.¹¹ One larger study conducted in urban and rural surveillance locations in Northern, Southern, Eastern and Western/Central India identified an OR of 3.0 (95% CI 2.5 to 3.7) for self-reported diabetes for those with graduate-level education versus those without formal schooling.⁶ Importantly, these studies were limited to selected geographical areas or cities in India. Our study has added to this literature

using a national population health survey with good coverage in rural areas across India.

Previous research in India has identified a strong positive relationship between SES and BMI among women and men in India.^{53–55} These studies are important because they have used similar markers of SES in the Indian context along with an objectively defined outcome (height and weight were measured in NFHS and not self-reported). BMI (along with other measures of body weight) is an important risk factor for the development of type 2 diabetes.^{47 49 56} Therefore, the consistency of our findings of a positive SES–diabetes association after controlling for BMI is encouraging. If BMI is part of the causal pathway between SES and diabetes, attenuation in the effect size for markers of SES would be expected. The graded and positive relation between household wealth and diabetes after accounting for BMI suggests that there are additional effects of household wealth on diabetes that are not mediated by BMI. The effects of social caste and education were largely attenuated after the inclusion of household wealth and prior to the inclusion of BMI. Household wealth was the strongest socioeconomic factor associated with self-reported diabetes, suggesting that social and behavioural changes associated with diabetes in India may be more closely related to increasing wealth and/or standard of living than educational attainment.

When compared with other studies in India, the overall prevalence of diabetes in the NFHS-3 was not high. This may have resulted from a combination of using self-reports of diabetes, the younger age of the NFHS-3 target population and sampling from the general population that included a high proportion of respondents in rural areas. Among individuals over 30 years of age, the prevalence was 2.5% (3.0% in men and 2.2% in women). Other studies using in rural India using similar age groups and blood measurements have reported diabetes prevalence of 4%, and a study from rural Andhra Pradesh found a prevalence of 12% based on combination self-report and blood measurements.^{17 57}

The current national estimate for diabetes prevalence in India is about 7% of the adult population aged 20–79 years. This estimate is based on three relatively recent and larger scale studies using a combination of oral glucose tolerance testing and self-reports of diabetes.^{4–6} There continues to be considerable uncertainty in estimates of diabetes for the whole of India due to the limited study locations (with a focus on urban areas), wide variation in survey sampling methodology, differences in diabetes diagnostic criteria and age groups studied. These differences in study design have hindered direct comparison of the prevalence between studies, across regions and over time. The NFHS-3 provides an important benchmark because it is the first nationally representative survey of diabetes in India. Even if the prevalence estimates of diabetes have been underestimated in the NFHS-3, the observed SES–diabetes

associations are plausible and important. Previous studies have largely overlooked the importance of SES markers, which may be a key determinant of diabetes. Further large-scale population-based surveys can be strengthened by using simple finger-prick blood glucose measurements in addition to self-reports.

There has been considerable concern over the rising prevalence of diabetes in India, especially with studies on migrant Indian populations suggesting that South Asians may be more susceptible to the disease. In light of current findings, it appears that, at present, the more well-off segments of the Indian population are at greatest risk. This poses concerns on how to appropriately balance priorities to address the disease burden that afflicts the non-poor versus the poor in the context of India where >40% of the population continue to live in extreme poverty on <\$1.25/day.⁵⁸

Contributors SVS and DJC planned the study. DJC conducted statistical analyses and drafted the manuscript with supervision from SVS. Both authors participated in interpretation of the results and critical revisions of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The NFHS data are available through the Measure DHS project at <http://www.measuredhs.com/>

REFERENCES

1. Rao KS, Mukherjee NR, Rao KV. A survey of diabetes mellitus in a rural population of India. *Diabetes* 1972;21:1192–6.
2. Gupta OP, Joshi MH, Dave SK. Prevalence of diabetes in India. *Adv Metab Disord* 1978;9:147–65.
3. Ahuja MM. Recent contributions to the epidemiology of diabetes mellitus in India. *Int J Diab Developing Countries* 1991;11:5–9.
4. Ramachandran A, Snehalatha C, Kapur A, et al. High prevalence of diabetes and impaired glucose tolerance in India: National Urban Diabetes Survey. *Diabetologia* 2001;44:1094–101.
5. Sadikot SM, Nigam A, Das S, et al. The burden of diabetes and impaired glucose tolerance in India using the WHO 1999 criteria: prevalence of diabetes in India study (PODIS). *Diabetes Res Clin Pract* 2004;66:301–7.
6. Mohan V, Mathur P, Deepa R, et al. Urban rural differences in prevalence of self-reported diabetes in India—the WHO-ICMR Indian NCD risk factor surveillance. *Diabetes Res Clin Pract* 2008;80:159–68.
7. Singh RB, Sharma JP, Rastogi V, et al. Social class and coronary disease in rural population of north India. The Indian Social Class and Heart Survey. *Eur Heart J* 1997;18:588–95.
8. Singh RB, Niaz MA, Thakur AS, et al. Social class and coronary artery disease in a urban population of North India in the Indian Lifestyle and Heart Study. *Int J Cardiol* 1998;64:195–203.
9. Singh RB, Bajaj S, Niaz MA, et al. Prevalence of type 2 diabetes mellitus and risk of hypertension and coronary artery disease in rural and urban population with low rates of obesity. *Int J Cardiol* 1998;66:65–72.
10. Singh RB, Beegom R, Mehta AS, et al. Social class, coronary risk factors and undernutrition, a double burden of diseases, in women during transition, in five Indian cities. *Int J Cardiol* 1999;69:139–47.
11. Ramachandran A, Snehalatha C, Vijay V, et al. Impact of poverty on the prevalence of diabetes and its complications in urban southern India. *Diabet Med* 2002;19:130–5.
12. Gupta R, Gupta VP, Sarna M, et al. Serial epidemiological surveys in an urban Indian population demonstrate increasing coronary risk factors among the lower socioeconomic strata. *J Assoc Physicians India* 2003;51:470–7.
13. Reddy KS, Prabhakaran D, Jeemon P, et al. Educational status and cardiovascular risk profile in Indians. *Proc Natl Acad Sci U S A* 2007;104:16263–8.

14. Ajay VS, Prabhakaran D, Jeemon P, *et al*. Prevalence and determinants of diabetes mellitus in the Indian industrial population. *Diabet Med* 2008;25:1187–94.
15. Vijayakumar G, Arun R, Kutty VR. High prevalence of type 2 diabetes mellitus and other metabolic disorders in rural Central Kerala. *J Assoc Physicians India* 2009;57:563–7.
16. Gupta R, Kaul V, Agrawal A, *et al*. Cardiovascular risk according to educational status in India. *Prev Med* 2010;51:408–11.
17. Kinra S, Bowen LJ, Lyngdoh T, *et al*. Sociodemographic patterning of non-communicable disease risk factors in rural India: a cross sectional study. *BMJ* 2010;341:c4974.
18. Samuel P, Antonisamy B, Raghupathy P, *et al*. Socio-economic status and cardiovascular risk factors in rural and urban areas of Vellore, Tamilnadu, South India. *Int J Epidemiol*. Published Online First: 24 February 2012.
19. Zaman MJ, Patel A, Jan S, *et al*. Socio-economic distribution of cardiovascular risk factors and knowledge in rural India. *Int J Epidemiol*. Published Online First: 21 May 2012.
20. Jeemon P, Reddy KS. Social determinants of cardiovascular disease outcomes in Indians. *Indian J Med Res* 2010;132:617–22.
21. Beaglehole R, Bonita R, Alleyne G, *et al*. UN high-level meeting on non-communicable diseases: addressing four questions. *Lancet* 2011;378:449–55.
22. World Health Organization. *Prevention of Diabetes Mellitus. Report of a WHO Study Group. WHO Technical Report Series 844*. Geneva: World Health Organization, 1994.
23. Hu FB, Manson JE, Stampfer MJ, *et al*. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001;345:790–7.
24. Psaltopoulou T, Ilias I, Alevizaki M. The role of diet and lifestyle in primary, secondary, and tertiary diabetes prevention: a review of meta-analyses. *Rev Diabet Stud* 2010;7:26–35.
25. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health Aff (Millwood)* 2002;21:60–76.
26. Brown AF, Ettner SL, Piette J, *et al*. Socioeconomic position and health among persons with diabetes mellitus: a conceptual framework and review of the literature. *Epidemiol Rev* 2004;26:63–77.
27. Sims M, Diez Roux AV, Boykin S, *et al*. The socioeconomic gradient of diabetes prevalence, awareness, treatment, and control among African Americans in the Jackson Heart Study. *Ann Epidemiol* 2011;21:892–8.
28. Krishnan S, Cozier YC, Rosenberg L, *et al*. Socioeconomic status and incidence of type 2 diabetes: results from the Black Women's Health Study. *Am J Epidemiol* 2010;171:564–70.
29. Robbins JM, Vaccarino V, Zhang H, *et al*. Socioeconomic status and type 2 diabetes in African American and non-Hispanic white women and men: evidence from the Third National Health and Nutrition Examination Survey. *Am J Public Health* 2001;91:76–83.
30. Dinca-Panaitescu S, Dinca-Panaitescu M, Bryant T, *et al*. Diabetes prevalence and income: results of the Canadian Community Health Survey. *Health Policy* 2011;99:116–23.
31. Agardh E, Allebeck P, Hallqvist J, *et al*. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int J Epidemiol* 2011;40:804–18.
32. Mohan V, Sandeep S, Deepa R, *et al*. Epidemiology of type 2 diabetes: Indian scenario. *Indian J Med Res* 2007;125:217–30.
33. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 2010;87:4–14.
34. Mohan V, Deepa M, Deepa R, *et al*. Secular trends in the prevalence of diabetes and impaired glucose tolerance in urban South India—the Chennai Urban Rural Epidemiology Study (CURES-17). *Diabetologia* 2006;49:1175–8.
35. Ramachandran A, Snehalatha C, Latha E, *et al*. Rising prevalence of NIDDM in an urban population in India. *Diabetologia* 1997;40:232–7.
36. International Institute for Population Sciences (IIPS) and Macro International. *National Family Health Survey (NFHS-3), 2005–06: India: Volume I*. Mumbai: IIPS, 2007.
37. Macro International Incorporated. *Sampling Manual. Calverton, Maryland: DHS-III Basic Documentation No. 6*. 1996.
38. ICF Macro. *Demographic and Health Survey Interviewer's Manual. MEASURE DHS Basic Documentation No. 2*. Calverton, Maryland, USA: ICF Macro, 2011.
39. Chitnis S. Definition of the terms scheduled castes and scheduled tribes: a crisis of ambivalence. In: Pai Panandiker VA, ed. *The Politics of Backwardness: Reservation Policy in India*. New Delhi, India: Centre for Policy Research, 1997.
40. Rutstein SO, Johnson K. *The DHS Wealth Index. DHS Comparative Reports No. 6*. Calverton, Maryland, USA: ORC Macro, 2004.
41. Filmer D, Pritchett L. Estimating wealth effects without expenditure data-or tears: an application to education enrollments in states of India. *Demography* 2001;37:115–74.
42. World Health Organization (WHO) Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;363:157–63.
43. Subramanian SV, Jones K, Duncan C. Multilevel methods for public health research. In: Kawachi I, Berkman LF, eds. *Neighborhoods and Health*. New York, NY: Oxford University Press, 2003:65–111.
44. Goldstein H. *Multilevel Statistical Models*. 3rd edn. London: Arnold, 2003.
45. Rasbash J, Browne WJ, Healy M, *et al*. *MLwiN Version 2.20*. Bristol, UK: Centre for Multilevel Modelling, University of Bristol, 2010.
46. Browne WJ. *MCMC Estimation in MLwiN*. Bristol, UK: Centre for Multilevel Modelling, University of Bristol, 2009.
47. Koh-Banerjee P, Wang Y, Hu FB, *et al*. Changes in body weight and body fat distribution as risk factors for clinical diabetes in US men. *Am J Epidemiol* 2004;159:1150–9.
48. Vazquez G, Duval S, Jacobs DR Jr, *et al*. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev* 2007;29:115–28.
49. Guh DP, Zhang W, Bansback N, *et al*. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* 2009;9:88.
50. Okura Y, Urban LH, Mahoney DW, *et al*. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57:1096–103.
51. Subramanian SV, Subramanyam MA, Selvaraj S, *et al*. Are self-reports of health and morbidities in developing countries misleading? Evidence from India. *Soc Sci Med* 2009;68:260–5.
52. Registrar General of India. *Census of India 2001: C-13 Single Year Age Returns by Residence and Sex*. 2001. http://www.censusindia.gov.in/Tables_Published/C-Series/c_series_tables_2001.aspx
53. Subramanian SV, Smith GD. Patterns, distribution, and determinants of under- and overnutrition: a population-based study of women in India. *Am J Clin Nutr* 2006;84:633–40.
54. Ackerson LK, Kawachi I, Barbeau EM, *et al*. Geography of underweight and overweight among women in India: a multilevel analysis of 3204 neighborhoods in 26 states. *Econ Hum Biol* 2008;6:264–80.
55. Subramanian SV, Perkins JM, Khan KT. Do burdens of underweight and overweight coexist among lower socioeconomic groups in India? *Am J Clin Nutr* 2009;90:369–76.
56. Colditz GA, Willett WC, Rotnitzky A, *et al*. Weight gain as a risk factor for clinical diabetes mellitus in women. *Ann Intern Med* 1995;122:481–6.
57. Chow CK, Raju PK, Raju R, *et al*. The prevalence and management of diabetes in rural India. *Diabetes Care* 2006;29:1717–18.
58. World Bank. *Poverty Headcount Ratio at \$1.25 a Day (PPP) (% of Population)*. 2012. <http://data.worldbank.org/indicator/SI.POV.DDAY?page=1>

STROBE Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Article: Association between socioeconomic status and diabetes in India

Authors: Daniel J Corsi, SV Subramanian

	Item No	Recommendation
Title and abstract	1	(a) Design cross-sectional study, listed in abstract (b) abstract and article summary (page 3)
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported, (abstract, page 3, page 4-5)
Objectives	3	State specific objectives, including any prespecified hypotheses (abstract, page 3, page 4-5)
Methods		
Study design	4	Present key elements of study design early in the paper (abstract, page 6)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection (abstract, page 5-7)
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants (page 6)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable (page 7)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group (page 7-8)
Bias	9	Describe any efforts to address potential sources of bias (page 8-9; page 12)
Study size	10	Explain how the study size was arrived at (page 7)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why (page 8)
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (page 8-9) (b) Describe any methods used to examine subgroups and interactions (page 12) (c) Explain how missing data were addressed (page 7) (d) If applicable, describe analytical methods taking account of sampling strategy (page 8) (e) Describe any sensitivity analyses (page 12)
Results		
Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (page 8) (b) Give reasons for non-participation at each stage (n/a) (c) Consider use of a flow diagram (n/a)
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (page 9-10, table 1) (b) Indicate number of participants with missing data for each variable of interest (page 7; page 12)
Outcome data	15*	Table 1
Main results	16DJC	(a) Tables 1 (unadjusted), 2 (age, gender, marital status, religion, place of residence adjusted; and fully adjusted model)

		(b) Report category boundaries when continuous variables were categorized (in methods, e.g. for BMI)
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period (n/a)
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses (Page 12)
Discussion		
Key results	18	Summarise key results with reference to study objectives (page 13)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias (page 13)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence (page 16)
Generalisability	21	Discuss the generalisability (external validity) of the study results (pages 14-15)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based (title page)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.