Oxford Open Economics, 2024, **3**, odae008
https://doi.org/10.1093/ooec/odae008
Advance access publication date 16 February 2024
Research Article

Education gradient in differential health reporting error among older adults in India

Anna Choi [1,*], Arnab K. Basu [2], Nancy H. Chau [3] and T.V. Sekher [4]

[1] Sejong University, Department of Public Administration, 209 Neungdong-ro, Gwangjin-gu, Seoul, South Korea
[2] Charles H. Dyson School of Applied Economics and Management, Cornell University & Institute for the Study of Labor (IZA), Warren Hall, 137 Reservoir Ave, Ithaca, NY 14853, USA
[3] Charles H. Dyson School of Applied Economics and Management, Warren Hall, 137 Reservoir Ave, Cornell University, Ithaca, NY 14853, USA
[4] Centre for Ageing Studies, International Institute for Population Sciences (IIPS), Govandi Station Road, Deonar, Mumbai 400 088. INDIA.
*Correspondence address. Sejong University, Department of Public Administration, 209 Neungdong-ro, Gwangjin-gu, Seoul, South Korea E-mail: anna.choi@sejong.ac.kr

ABSTRACT

This paper studies the education gradient associated with health reporting errors for two highly prevalent non-communicable diseases among older adults in India. We analyze a novel data set—the Longitudinal Aging Study in India (2017–18) panel survey—to unpack the sources of health reporting error in a developing-country context for the first time. Our analysis points to a statistically significant level of false negative reporting (or over-reporting) for both high blood pressure and depression by those with no schooling and less than middle school education relative to their more highly educated peers. Interestingly, this result is driven by those whose income is above the poverty line. We further find false negative reporting for high blood pressure to be more prevalent among women with no schooling.

Key words: high blood pressure; depression; reporting error; education

Introduction

Target 3.4 of the 2030 Sustainable Development Goals (SDGs) adopted by all UN member countries benchmarks a reduction in premature mortality from non-communicable diseases by one-third through prevention and treatment, and an improvement in mental health and well-being. The purpose of these targets is to set supply-side guidelines for governments across the globe to invest resources in critical areas of healthcare delivery, including those for the betterment of mental health. After all, government spending on health and health-related activities in the developing world can be notoriously low – around 3% of gross domestic product (GDP) for the South Asian group of countries (World Bank, 2019)[1] – and the targets provided by the SDGs act as spending goalposts in key areas.

Accompanying the supply-side push, several demand-side interventions have been introduced in developing countries to raise health awareness and effect proactive health-seeking behaviour. These include campaigns to change behavioural attitudes toward cleanliness and disease prevention via inducing more handwashing, smoking cessation campaigns, promoting healthy eating habits and instilling the virtues of regular physical activity. However, bottlenecks are plenty – from a rise in chronic non-communicable diseases such as hypertension and diabetes (older adults suffer more from chronic rather than acute illnesses) to income traps (low income associated with lower ability to access preventative care) to gender bias in health-seeking behaviour (women are less likely to seek medical care especially where the majority of the doctors are male) and societal norms and mistrust of modern medicine (reliance on traditional doctors and vaccine hesitancy).[2]

Demand- or supply-side interventions notwithstanding, health-seeking behaviour is guided by the awareness of one's health status and therein lies the contribution of this paper. Leveraging a unique survey containing the self-reported and objectively measured health conditions of adults aged 45 years or older in India, we contribute to the literature on differential health reporting error by education levels in a developing-country setting.[3] The Longitudinal Aging Study in India (LASI) conducted over 2017–18 catalogues self-reports and objective measures of high blood pressure, depression and physical activity, allowing for unique insights into the magnitude of reporting error for a widely prevalent non-communicable disease – hypertension – and a neglected health condition in developing countries – depression. Notable here is the fact that LASI is the only nationally

[1] For North America, health spending was roughly 16%, and for OECD countries, health spending was 12.5% of GDP in 2019 (World Bank, 2019).

[2] For an analysis of the success of smoking cessation campaigns in South Asia see Iqbal *et al.* (2022), for an analysis a large-scale handwashing intervention and its impact on respiratory illness in Bangladesh see Najnin *et al.* (2019), for inadequate number of facilities to treat chronic non-communicable diseases in India see Dey *et al.* (2012), how the low availability of women physicians in rural areas of India is negatively affecting health-seeking behaviour see Bhan *et al.* (2020) and for the origins of mistrust of modern vaccines in Africa see Lowes & Montero (2021).

[3] In effect, our analysis here follows the methodology of Choi & Cawley (2018), undertaken for the USA.

Received: September 12, 2023. **Revised:** February 6, 2024. **Accepted:** February 7, 2024

representative dataset available from a developing country that captures both self-reported and objective measures of depression.

While studies on differential health reporting error are more common in developed countries, data paucity makes such analysis impossible to conduct for developing countries. Our paper is thus unique in this regard. This is also the first paper to our knowledge to investigate the differential reporting error by education levels in self-reported depression diagnosis compared with depressive symptoms as measured by the Center for Epidemiological Studies-Depression (CES-D) scores. In what follows, we discuss in turn the importance of analyzing the health–education gradient, the magnitude of high blood pressure and depression among the Indian population, and earlier research conducted via the use of the LASI pilot study in 2010.

Education gradient in Health Status Self-Reports

There are at least three pathways through which the reporting of health status and education can be related. First, health and education may be causally related – higher levels of education lead to better health. Second, selection may occur whereby the healthy become better educated. And finally, health and education may both be correlated with a confounder. For example, good health later in life and high educational attainment are both a result of early life interventions such as better nutrition and vaccinations. Our focus is on the selection problem given the cross-section data at our disposal, which precludes establishing a causal relationship between educational attainment and health. In addition, variables such as income are endogenous with current health status whereas education, completed several years prior for adults aged 45 years or older, is not. It is thus unsurprising that the health–education gradient is a popular subject for research. Several studies based in developed countries, primarily the USA, have established that adults with higher educational attainment have better health and lifespans. For instance, Grossman (2008) finds that education (in terms of years of formal schooling) is one of the most important positive correlates of health. Different mechanisms and factors could drive this relationship. Cutler and Lleras-Muney (2010) find that income, health insurance and family background account for 30%, knowledge and cognitive ability measures account for another 30% and social networks account for 10% of the relationship between education and health.

However, most of these studies that explore the education gradient associated with health status self-reports rely on self-reported health measures, which could be prone to various biases and errors. It is possible that the better educated are more accurate in reporting their health conditions and behaviours, as evidenced in the studies undertaken by Choi and Cawley (2018) and Johnston *et al.* (2009). Social desirability bias may also influence how individuals respond to survey questions on health status or behaviours, especially when a more socially acceptable response is deemed to be expected. It is also possible that education is correlated with other skills and knowledge, such as health literacy. Van der Heide *et al.* (2013) find evidence of health literacy possibly mediating poor self-reported health status among those with lower levels of schooling.

For our geographical context, India, studies on the health status reporting–education gradient are sparse. Cramm *et al.* (2015) find low education levels to be a significant predictor of smoking status and lower levels of physical activity. Raghupathi and Raghupathi (2020) find that tertiary education is critical in positively influencing infant mortality, life expectancy, child vaccination and healthcare enrollment rates.

High blood pressure and depression in India

The annual number of deaths from cardiovascular disease (CVD) in India was projected to rise from 2.26 million (1990) to 4.77 million (2020), according to Huffman *et al.* (2011). This is not surprising given that India has one of the highest burdens of CVD worldwide. Prevalence rates of CVD have been estimated to have ranged from 1.6% to 7.4% in rural populations and from 1% to 13.2% in urban populations over the past several decades (Huffman *et al.*, 2011). Geldsetzer *et al.* (2018) estimate an even higher prevalence of hypertension in the urban areas of India, and compared with the poorest quintile, the richest quintile had only a moderately higher probability of being diagnosed with hypertension. Interestingly, Geldsetzer *et al.* (2018) do not find any difference in the probability of hypertension across educational attainment. However, Pednekar *et al.* (2011) find that CVD mortality is significantly greater among Indian men with low educational status, while the association is not clear in women.

The prevalence of depression, and even more so its causes and consequences, is an under- researched area, particularly in a developing-country context. Two recent studies by Bernadine (2021) and Dandona *et al.* (2020) offer some key insights into the extent of the issue. Bernadine (2021) reiterates the World Health Organization (WHO) finding that the economic loss associated with the prevalence of mental health disorders in India is extremely high, at US$1.03 trillion. In addition, data from the National Mental Health Survey (NMHS, 2015–16) point to mental health disorders disproportionately affecting households with lower income, less education and underemployment (cited in Bernadine (2021)). Furthermore, median out-of-pocket expenditure by families on treatment and travel to access care amount to 1000–1500 Indian rupees (INR) per month, equivalent to ~US$13–20: a large sum for poorer families. Dandona *et al.* (2020) estimate that 45.7 million people had depressive disorders in India in 2017. The highest prevalence was observed in the Southern Indian states of Tamil Nadu, Kerala, Goa and Telangana (which rank high on the Social Development Index/SDI), Andhra Pradesh (which ranks among the middle SDI group of Indian states) and Odisha (which ranks among the low SDI group of Indian states). Dandona *et al.* (2020) further find no difference in overall prevalence between men and women and that depressive disorders increase with age, with the highest prevalence observed in older adults, especially among women starting at 45 years. Neither of these two studies, however, analyze the prevalence of mental disorders by educational attainment or income level.

Among South Asian countries, data between 1975 and 2015 show the incidence of high blood pressure being on the rise in all countries in the region (Afghanistan, Nepal, Pakistan, Bangladesh, India, Bhutan and Sri Lanka). Amongst males, 26.6% of the Indian population suffering from high blood pressure in 2015 compared with 31.5% in Pakistan and 24.5% in Bangladesh. Amongst females, India, Pakistan and Bangladesh had comparable prevalences at 24.7%, 29.5% and 24.9%, respectively, in 2015.[4] For depression, the World Population Review[5] reports 4.5% of the Indian population suffering from the disease as compared with 4.2% in Pakistan and 4.1% in Bangladesh in 2017.

[4] See NCD-RisC (Risk Factor Collaboration): https://ncdrisc.org/blood-pressure-raised-ranking.html

[5] https://worldpopulationreview.com/country-rankings/depression-rates-by-country

Table 1: Descriptive statistics

Variable	Overall N	Mean	Standard deviation
No education/schooling	64 866	50.40%	0.5
Less than middle school	64 866	23.00%	0.42
Middle or high school education	64 866	20.90%	0.41
College education	64 866	3.60%	0.19
Graduate education	64 866	2.10%	0.14
Age	64 866	60.17	10.68
Male	64 866	45.95%	0.5
Married	64 662	92.70%	0.26
Hospitalized in the past year	63 375	9.10%	0.44
Insurance coverage	64 211	20.30%	0.4
Visited a healthcare facility in the past year	64 245	72.20%	0.45
Current work status	64 851	46.70%	0.5
Annual earnings (INR)	64 204	41688.40	89627.14
% With high blood pressure (self-report)	64 689	27.5	0.45
% With high blood pressure (measured)	59 629	42.25%	0.49
% With depression (self-report)	64 673	0.6%	0.077
% With depression (based on CESD-10 scores)	63 807	46.04%	0.498

Notes 1. Summary statistics based on LASI Wave I (weighted) (IIPS, 2023). 2. Includes adults 45 and older; excludes responses if reported by a proxy or had missing values for the education variable.

Research using LASI pilot study (2010) and Wave 1 (2017–18)

The pilot Longitudinal Aging Study (LASI) was undertaken in 2010, and a few studies based on this survey have been published. Cramm et al. (2015) is the first such study to find that subjective (compared with objective) health measures underestimate the health status of older adults, and, as mentioned earlier, also finds a negative relationship between educational attainment and both smoking status and physical activity. Vellakkal et al. (2013) find that self-reports (compared with objective measures) underestimate the prevalence of hypertension among the poor, while Lee et al. (2015) analyze the prevalence of hypertension using data from LASI Wave 1 (2017–18) and the pilot study (2010) and find that almost 46% of adults aged 45 or older are hypertensive. Of those diagnosed with hypertension, 39% take anti-hypertensive medication and access to public healthcare is a key predictor of hypertension treatment among the lowest income group. Onur and Velamuri (2018) find that measurement error in self-reported hypertension and lung disease is over 80%—larger than what is found in high-income countries and substantially understates the true disease burden for both. While our paper also focuses on high blood pressure/hypertension, we depart on two fronts, beyond the obvious focus on depression as opposed to lung disease. First, Onur and Velamuri uses LASI Pilot data from four Indian states (Punjab, Rajasthan, Kerala and Karnataka) and analyzes only incidences of false-negative reporting (reporting not having the disease but testing positive for it). We instead use the first wave of a nationally representative LASI sample from 2017 to 2018 and analyze false negative, false positive (reporting having the disease but testing negative for it) and accuracy (reporting positive/negative for a disease and testing also positive/negative for it). Second, Onur and Velamuri focuses on lack of access to medical facilities, high levels of health illiteracy, low rates of health insurance, poverty and lack of equity in the delivery of health services as the confounders for false negative reporting of both hypertension and lung disease, while we focus on the education gradient after controlling for income, access to healthcare, physical activity recall memory and family history of hypertension and depression.

With the background in place, we discuss our descriptive statistics, empirical strategy, results and robustness checks in the next sections.

Data and descriptive statistics

We use the first wave (2017–18) data from the LASI, which contains self-reported and objective measures of high blood pressure and depression. The LASI contains a nationally and state representative sample of adults aged 45 and older in India and covers 30 states and 6 Union Territories of India (the Sikkim survey was undertaken in 2021).[6] Given the widespread prevalence of CVD in India and the paucity of studies on depression in developing countries, we focus on the education gradient of reporting error of these two health conditions.

In the survey, respondents report whether any health professional has ever diagnosed them with high blood pressure and any neurological or psychiatric problems such as depression.[7] High blood pressure is defined as self-reported physician diagnosis or elevated blood pressure (BP) on measurement of systolic BP $\geq$ 140 mmHg or diastolic BP $\geq$ 90 mmHg. We consider the respondent to have high blood pressure with the following objective measurements: average systolic 140 mmHg or higher, and average diastolic 90 mmHg or higher.

For depression, respondents' answers on the 10-item Center for Epidemiological Studies-Depression (CES-D) questionnaire are used. The CES-D is a standard evaluation tool that measures a respondent's mental health status over the past week. The scores from the responses can serve as a close proxy for depression screening and have good predictive accuracy when compared with the 20-item scale among older adults (Andresen et al., 1994; Gellis, 2010). If the respondent has a score of 10 or above, we consider them to have depressive symptoms (Radloff, 1977;

[6] Perianayagam et al. (2022) provides a comprehensive overview of the cohorts surveyed in LASI Wave 1 (2017–18), while Bloom et al. (2021) explains how the LASI survey can advance ageing research in response to the challenges posed by population ageing.

[7] The questionnaire wording for high blood pressure is: 'Who first diagnosed you with high blood pressure or hypertension? 1. A doctor (MBBS degree), 2. Ayurvedic/Unani/Homeopathic/Siddha, 3. Other, please specify.' In our analysis, we only account for those diagnosed by 1 (doctor) or 2 (Ayurvedic/Unani/Homeopathic/Siddha). For depression, the questionnaire wording is: 'Has any health professional ever diagnosed you with any neurological, or psychiatric problems such as depression, Alzheimer's/dementia, unipolar/bipolar disorders, convulsions, and Parkinson's? 'There is a subsequent question that asks respondents to choose which neurological or psychiatric problem they were diagnosed with, including depression.

Andresen *et al.*, 1994; Gellis, 2010). Later in the paper, we conduct robustness checks using different thresholds of CES-D (25th, median, 75th and 90th percentile) to categorize someone as suffering from depression.

We next turn to the descriptive statistics. As Table 1 shows, after removing observations with missing data on education, we have 64,866 older adults (45 and older) in our analysis sample. Roughly 50% of respondents in our sample do not have any formal education, the mean age is 60 years, 45% are male and 91% are married. A large percentage (45%) are actively employed and only 20% have health insurance coverage. Nevertheless, a high percentage of respondents (72%) have visited a healthcare facility within the past year of the survey.

Our sample, constructed of only those respondents who report their education levels, compares well with the overall LASI sample as reported by IIPS *et al.* (2020). According to the IIPS study, across the 45–59 age group, 13% of men and 9% of women have less than primary school education, 15% of men and 11% of women have completed primary schooling, 23% of men and 13% of women have completed secondary schooling, 7% of men and 3% of women have completed higher secondary schooling, while 9% of men and 4% of women have completed college or higher degrees. Across the 60+ age group, the percentage distribution of educational attainment is similar, with 15% of men and 9% of women having less than primary school education, 15% of men and 8% of women having completed primary schooling, 21% of men and 8% of women having completed secondary schooling, 5% of men and 1% of women having completed higher secondary schooling and 5% of men and 1% of women having completed college or higher degrees. Notably, 34% of men and 61% of women in the overall LASI sample (45 years and older) have no schooling. Our sample also compares favourably with the overall 45+ LASI sample in terms of the demographic characteristics of the respondents (IIPS *et al.*, 2020) with 54% of our sample consisting of adults in the 45–59 age group and 46% of adults aged 60 and over, and 42% of the sample aged 45 and over are men, while 58% are women, 68% reside in rural areas, while 32% reside in urban areas.

Measures of reporting error

We construct three dichotomous measures of reporting error:

- Accuracy: If the self-reports of health condition or behaviours match objective measurements, then we code the respondent as providing an accurate report (equal to 1 if this is the case and 0 otherwise). For example, if a respondent reports having high blood pressure and the objective measurements also indicate high blood pressure based on the readings, we code this response as an accurate self-report of high blood pressure. On the other hand, if the respondent mentions not having been diagnosed with high blood pressure but objective measurements show otherwise, we code the response as 0 for accurate reporting of high blood pressure. If a respondent is taking medication for high blood pressure, it is possible that they will have blood pressure measurements lower than the threshold for hypertension. We take this into account and code respondents who report taking medication for high blood pressure but have blood pressure measurements lower than the threshold as providing accurate self-reports of high blood pressure. We take the same approach for those respondents who report that they currently take anti-depressants, which could influence the CES-D scores.
- False-negative: When respondents report not having high blood pressure/depression but the blood pressure reading/CES-D score shows otherwise, the respondent is coded as 1 (0 otherwise). For example, when a respondent reports not having been diagnosed with depression, but the CES-D score shows otherwise, the respondent is coded as providing a false negative report of depression.
- False-positive: When respondents report having high blood pressure/depression but the blood pressure readings/CES-D score indicate otherwise, the respondent is coded as 1 (0 otherwise). For example, when a respondent report having high blood pressure, but the blood pressure readings indicate otherwise, the respondent is coded as providing a false positive report of high blood pressure. The matrix below provides a visual characterization of the various types of reporting discussed above.

	Objective negative	**Objective positive**
Subjective negative	True negative	False negative
Subjective positive	False positive	True positive

Note that 'Accuracy' of reporting is captured along the diagonal of the matrix[8] while the cross-diagonals capture the two types of reporting errors we focus on in this paper. For our overall sample, 74.8% of older adults accurately report high blood pressure, while 56.9% of older adults accurately report depression, and 46.1% false negatively reported high blood pressure and 99% false negatively reported depression,[9] and 8.3% false positively reported high blood pressure and 0.3% false positively reported depression.

Figure 1 shows the gap between self-reported and objectively measured incidence of high blood pressure and depression by education levels in our sample. A rough comparison shows that, on average, the gap between self-reported and objectively measured high blood pressure is 10–15% for all education levels except those with graduate education, where the gap is around 20%. However, for depression, the gap between self-reported and objectively measured incidence by education levels is reversed. Self-reporting is low across the board, but the self-reported and objectively measured gap is greatest for those with no education, at around 50%, and lowest for those with graduate education, at around 30%.

Figure 2 shows the percentage of older adults who accurately predict high blood pressure and depression. Almost 75% of older adults with no education report high blood pressure accurately; the figure is consistently this high or higher for all education levels, reaching 79% for those with a bachelor's degree, except the highest – graduate education – for which it is around 66%. We

[8] 'True positive' (self-reported sick and correctly identified objectively) and 'true negative' (self- reported healthy and correctly identified objectively) measures are identical to the accuracy measure.

[9] A meta-analysis of different studies showed that up to about 50% of depression cases are not detected by general practitioners and successful detection of depression was less successful among older adults than younger adults (Carey *et al.*, 2015). There are several reasons cited in the literature for under-diagnosis and underreporting (i.e. not reporting depression in self-reported questionnaire) of depression among older adults. Stigma is cited as a barrier in recognizing depression (Smith and Meeks, 2019; Devita *et al.*, 2022). Depression among older adults could show various symptoms (fatigue, pain, insomnia, etc.) which could be similar to symptoms of other illnesses such as dementia (Devita *et al.*, 2022). It is possible that older adults with less severe depression symptoms consider their symptoms as a response to day to day life stress (Birrer and Vemuri, 2004). There are different scales such as geriatric depression scale to screen for depression specifically among older adults (Krishnamoorthy *et al.*, 2020), which might yield different results with respect to false negative or positive reporting of depression.

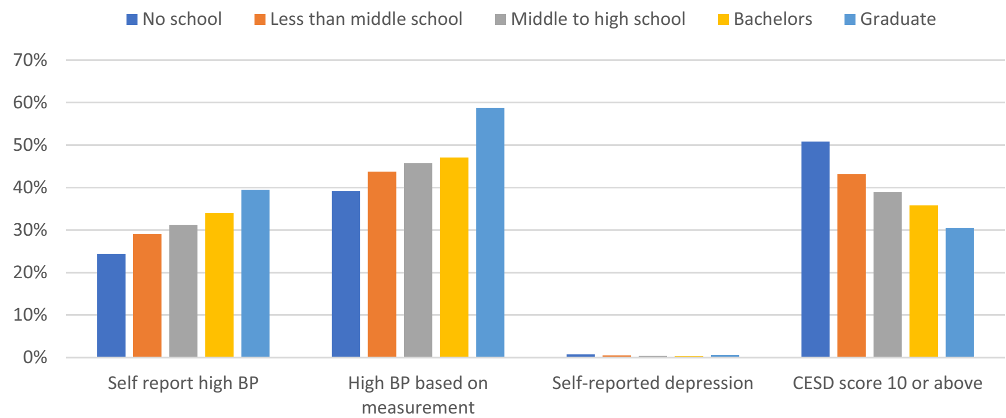

Figure 1: Self-reported and objectively measured blood pressure and depression by education levels (%)

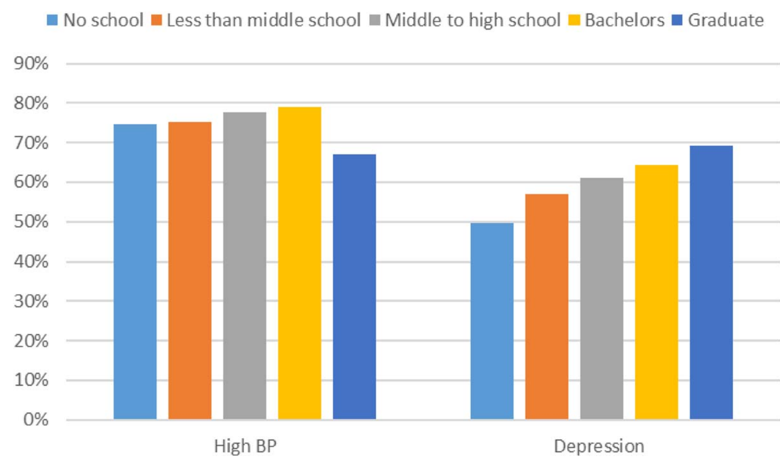

Figure 2: Accuracy of reporting of high blood pressure and mental depression by education levels (%)

observe the opposite pattern with depression: the least educated group have the lowest percentage of accurate reporting (49%) and the most educated group have the highest (69%).

Methods

We estimate linear probability regression models for the three types of reporting (accurate, false negative and false positive) as the dependent variables. In the reported marginal effects in the tables that follow, we use individual-level weights to account for the multistage stratified area probability cluster sampling design with stage stratification and sample selection stages.[10] The main specification is as follows:

$$\text{Reporting Error}_i = \alpha + \beta \text{Education}_i + \gamma X_i + \delta_{state} + \varepsilon_i. \quad (1)$$

The main explanatory variable is the level of education. We create an indicator variable for no schooling or education, less than middle school, middle or high school graduate, college graduate and post-graduate degree. About half of the sample report no schooling or education (Table 1). The share of older adults reporting having a college or graduate degree is less than 10%. The omitted reference category is middle or high school graduates. We also control for correct recall of words as a proxy for cognitive ability: after the interviewer read a set of 10 words, respondents were asked to recall as many as possible in 2 minutes. On average, respondents recalled about five words correctly. Number of words recalled is negatively associated with no schooling or education but positively associated with increasing levels of education. X is a vector of individual characteristics including gender, age, marital status, first language, religion and whether the respondent lives in a rural area. We also include state fixed effects (δ_{state}) in our baseline model.

[10] We exclude from the sample those who responded via a proxy rather than themselves, and those with missing values for education. Note that we also report the results from Probit regressions for comparison with our LPM in the Appendix. Appendix Tables 1–9 report the Probit estimates that mirror corresponding LPM estimates in Tables 2–12 in the paper.

Results

As Table 2 shows, older adults with no schooling or with less than middle school education report high blood pressure and depression less accurately than those with middle or high school education. This remains true (with smaller point estimates) when we add different control variables such as earnings, healthcare utilization and community-based identifiers such as scheduled caste later in our robustness checks. Those with no schooling are 3 percentage points less likely to accurately report high blood pressure and 5 percentage points less likely to accurately report depression than those with middle or high school education. Similarly, those with less than middle school education are 2 percentage points less likely to accurately report high blood pressure and 3.2 percentage points less likely to accurately report depression than those with middle to high school education.

Table 2: Linear probability regression results for accurate, false negative and false positive reporting of health conditions and behaviours (overall sample)

Variables	Accurate reporting		False negative reporting		False positive reporting	
	High blood pressure	Mental depression	High blood pressure	Mental depression	High blood pressure	Mental depression
No education	−0.030**	−0.050***	0.153***	0.001	−0.005	0.001
	(0.013)	(0.013)	(0.020)	(0.002)	(0.007)	(0.001)
Less than middle school	−0.020*	−0.032**	0.060***	0.003	0.015*	0.001
	(0.012)	(0.013)	(0.018)	(0.003)	(0.008)	(0.001)
Bachelor's	0.014	0.011	−0.056*	0.004	0.010	0.000
	(0.019)	(0.025)	(0.029)	(0.004)	(0.015)	(0.002)
Graduate	−0.097	0.094*	0.033	−0.004	0.010	0.003
	(0.072)	(0.056)	(0.071)	(0.006)	(0.021)	(0.004)
Observations	59 128	63 251	26 398	27 444	32 730	36 035
F-test	2.22	4.38	23.22	0.84	3.26	0.43
Prob > *F*	0.0832	0.0044	0.0000	0.473	0.0206	0.7332

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the overall sample. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 3. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, and state fixed effects. 4. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 5. Standard errors in parentheses. 6. F-test results for equality of coefficients. 7. Source: authors' construction based on LASI Wave I (weighted).

Those with graduate education are 9.4 percentage points more likely to accurately report depression in the survey than those with middle and high school education only.

Below, Fig. 3 summarizes the estimates reported in Table 2 for ease of interpretation.

A closer look at reporting error (Table 2) shows false negative reporting to be dominant (where respondents over-report their health, e.g. stating an absence of high blood pressure in the self-report where the objective measurement indicates otherwise). As with the results on accurate reporting, the point estimates are larger for those with no education, who are more likely to give a false negative report of high blood pressure than those with middle or high school education. For reports of high blood pressure, we observe a monotonic decrease in the degree of false negative reporting based on education levels. College-educated respondents are 5.6 percentage points less likely to give a false negative report of high blood pressure than those with middle or high school education.

The results on false positive reporting of health suggest that for an asymptomatic condition such as high blood pressure, those with less than middle school are more likely to under-report their health (e.g. saying that they have high blood pressure when objective measurements do not support this) than those with middle or high school education**.** We should be cautious in interpreting this result given that we have a marginally significant coefficient and because there are very few false positive cases, possibly due to reporting error.

Robustness checks: income, healthcare utilization status and different thresholds for depression

We further estimate the main baseline model controlling for additional variables related to income and healthcare utilization, such as income levels including current work status and annual earnings (Tables 2 and 3), respondent's health insurance coverage, and healthcare utilization in the past year (whether respondents have any health insurance coverage, whether they visited a healthcare facility or were hospitalized in the past year) (Table 4). We control for these variables as they may influence the likelihood of discovering certain health conditions such as high blood pressure, which is asymptomatic.

Furthermore, we estimate the main specification with all the control variables used in the baseline model as well as whether the respondent belongs to a scheduled caste,[11] which may influence their health through eating habits or other health-seeking behaviours (Table 5). The findings are remarkably consistent when additional income- and health-related variables are added. Adding scheduled caste in addition to the income and health variables also does not change our baseline findings (Table 5). The magnitude of the point estimates is similar or slightly smaller than in our baseline model (Table 2), but the patterns are mostly the same. Those with no schooling are significantly less likely to accurately report both high blood pressure and depression. Similarly, the less educated are more likely to make false negative reports regarding high blood pressure, whereas those with a bachelor's degree are less likely to do so in each of our robustness checks (Tables 3–5).

We conduct additional sensitivity analysis for depression. We use the threshold of score 10 from the CES-D 10-item questionnaires to define someone as suffering from depression based on objective measurement. To check whether our results are sensitive to the threshold we use, we estimate the baseline model with additional controls (income- and health-related variables as in Tables 3 and 4) for depression using four different thresholds (the 25th percentile of the sample, which is CES-D score of 7 or higher; the 50th, score of 9 or higher; the 75th, score of 12 or higher and the 90th, score of 15 or higher). We re-create the three outcome variables for reporting error for depression separately based on these different thresholds. The main results remain robust, as shown in Table 6: the less educated are less likely to accurately report depression at all thresholds than those who have middle or high school education. Even at the 90th percentile, this remains true: the better educated are more likely to accurately report depression than those with middle or high school education.

[11] We control for whether the respondent belongs to a scheduled caste because of previous literature suggesting significant association between caste membership and health behaviours and mortality rates (Coelho *et al.*, 2016). Our results remain the same with a larger point estimate without this control.

Table 3: Linear probability regression results for accurate, false negative and false positive reporting of health condition (overall sample) with income-related controls

	Accurate reporting		False negative reporting		False positive reporting	
Variables	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**
No education	−0.029**	−0.052***	0.138***	−0.000	−0.003	0.001
	(0.013)	(0.014)	(0.020)	(0.002)	(0.007)	(0.001)
Less than middle school	−0.019	−0.032**	0.055***	0.003	0.015*	0.001
	(0.012)	(0.013)	(0.019)	(0.003)	(0.009)	(0.001)
Bachelor's	0.020	0.005	−0.054*	0.003	0.009	0.000
	(0.020)	(0.026)	(0.029)	(0.004)	(0.015)	(0.002)
Graduate	−0.079	0.077	0.017	−0.006	0.012	0.004
	(0.068)	(0.056)	(0.067)	(0.007)	(0.021)	(0.004)
Observations	58 595	62 675	26 211	27 246	32 384	35 654

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the overall sample. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, and state fixed effects, work status and annual earnings in rupees. 5. LPM indicates linear probability model. 6. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 7. Standard errors in parentheses. 8. Source: authors' construction based on LASI Wave I (weighted).

Table 4: Linear probability regression results for accurate, false negative and false positive reporting of health condition (overall sample) with health-related controls

	Accurate reporting		False negative reporting		False positive reporting	
Variables	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**
No education	−0.026**	−0.055***	0.140***	0.001	−0.006	0.001
	(0.013)	(0.013)	(0.018)	(0.002)	(0.007)	(0.001)
Less than middle school	−0.019*	−0.032**	0.064***	0.004	0.014	0.001
	(0.012)	(0.013)	(0.016)	(0.003)	(0.009)	(0.001)
Bachelor's	0.020	0.001	−0.076**	0.004	0.012	0.000
	(0.020)	(0.025)	(0.031)	(0.004)	(0.015)	(0.002)
Graduate	−0.089	0.080	−0.010	−0.005	0.011	0.003
	(0.069)	(0.055)	(0.058)	(0.006)	(0.022)	(0.004)
Observations	58 131	62 246	25 989	26 872	32 142	35 598

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the overall sample. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, and state fixed effects, whether respondent has insurance coverage, hospitalization in past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors' construction based on LASI Wave I (weighted).

Table 5: Linear probability regression results for accurate, false negative and false positive reporting of health condition (overall sample) with income, health and caste controls

	Accurate reporting		False negative reporting		False positive reporting	
Variables	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**
No education	−0.024*	−0.053***	0.124***	0.001	−0.003	0.001
	(0.013)	(0.013)	(0.019)	(0.002)	(0.007)	(0.001)
Less than middle school	−0.018	−0.030**	0.060***	0.004	0.014	0.001
	(0.012)	(0.013)	(0.017)	(0.003)	(0.009)	(0.001)
Bachelor's	0.028	−0.002	−0.071**	0.003	0.011	0.000
	(0.020)	(0.026)	(0.032)	(0.004)	(0.015)	(0.002)
Graduate	−0.070	0.068	−0.024	−0.007	0.014	0.004
	(0.064)	(0.053)	(0.056)	(0.007)	(0.022)	(0.004)
Observations	57 445	61 493	25 730	26 568	31 715	35 147

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the overall sample. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, scheduled caste, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors' construction based on LASI Wave I (weighted).

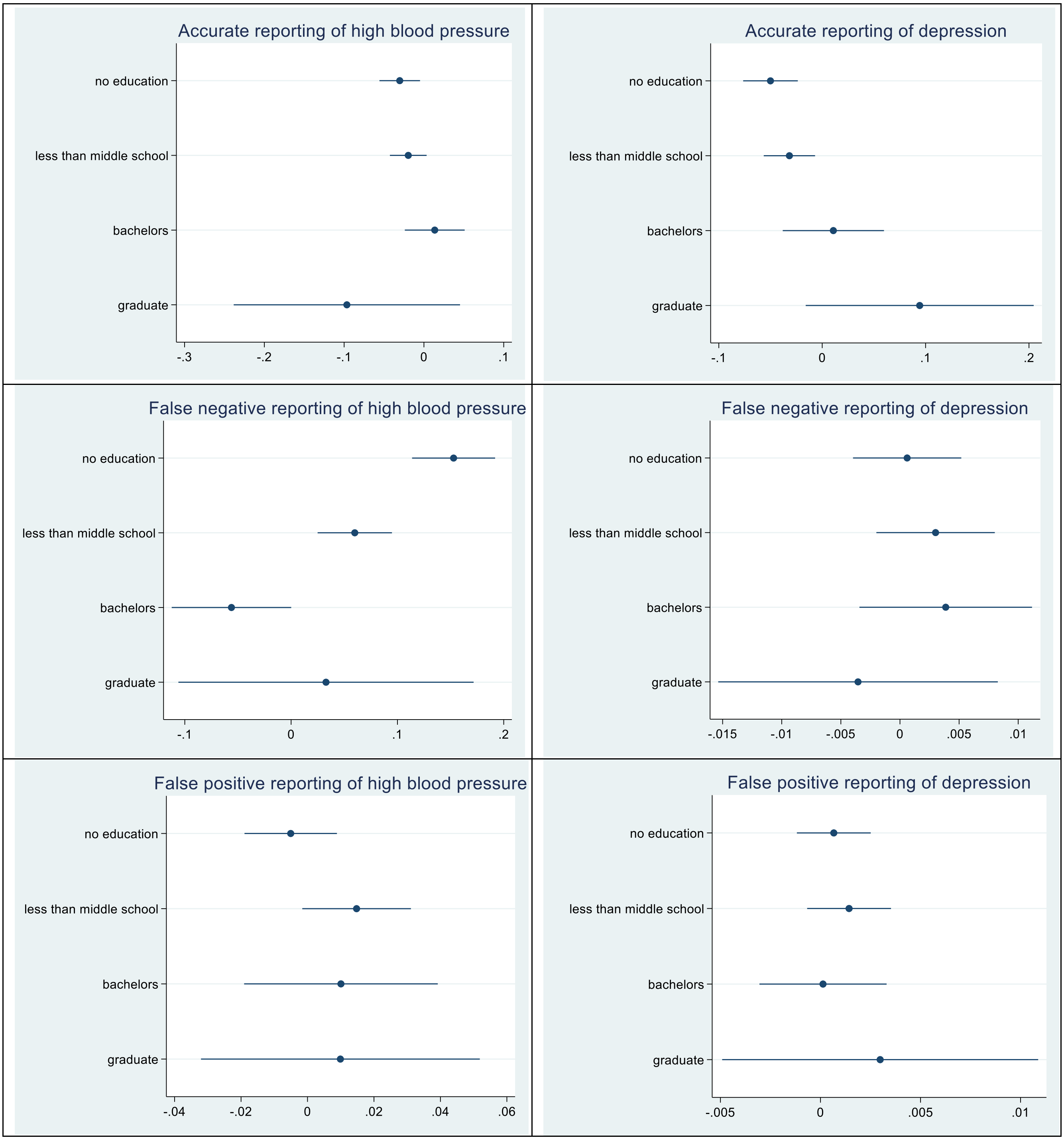

Figure 3: Visual presentation of estimates for accurate, false negative and false positive reporting of health conditions and behaviours (overall sample)

Subgroup analysis: gender, income, family medical history, physical activity and diagnosis recall

In addition to different robustness checks, we conduct several subgroup analysis based on categories such as gender, poverty status and whether the respondent's family has a history of high blood pressure. In Tables 7 and 8, we show the main results separately by gender, as studies have found a significant gender gap in health among adults in India (Lee *et al.*, 2015). Although the main results remain the same, this is primarily driven by men, as shown in Table 7. Among women (Table 8), although the direction of the coefficient is the same for results on accurate, false negative and false positive reporting of health, many coefficients are no longer significant even with a larger sample size than that of men. Women with no schooling are significantly more likely to make false negative reports regarding high blood pressure than those with middle or high school education. On the other hand, women with college education are significantly more likely to make false negative reports regarding depression than those with middle or high school education.

As noted in other studies that explore the income–health gradient in India, there is a significant difference in health conditions

Table 6: Linear probability regression results for accurate, false negative and false positive reporting of health condition (overall sample) with different CES-D thresholds

Accurate reporting				
	Percentiles			
	25th	50th	75th	90th
No education	−0.062***	−0.057***	−0.061***	−0.028***
	(0.012)	(0.013)	(0.011)	(0.007)
Less than middle school	−0.024**	−0.031**	−0.045***	−0.018***
	(0.012)	(0.012)	(0.010)	(0.007)
Bachelor's	0.035	0.010	0.013	0.018*
	(0.023)	(0.026)	(0.016)	(0.010)
Graduate	0.127**	0.104*	0.058**	0.028**
	(0.060)	(0.054)	(0.025)	(0.013)
Observations	61 682	61 682	61 682	61 682
False negative reporting				
	Percentiles			
	25th	50th	75th	90th
No education	−0.001	0.001	0.002	0.004
	(0.001)	(0.002)	(0.004)	(0.009)
Less than middle school	0.001	0.002	0.008**	0.012
	(0.002)	(0.002)	(0.004)	(0.009)
Bachelor's	−0.000	0.001	0.003	0.002
	(0.002)	(0.003)	(0.007)	(0.017)
Graduate	−0.004	−0.006	0.001	−0.026
	(0.004)	(0.005)	(0.012)	(0.032)
Observations	48 144	36 158	16 986	7581
False positive reporting				
	Percentiles			
	25th	50th	75th	90th
No education				
	−0.000	−0.000	0.001	0.001*
Less than middle school	(0.001)	(0.001)	(0.001)	(0.001)
	0.001	0.001	0.001*	0.001
Bachelor's	(0.001)	(0.001)	(0.001)	(0.001)
	−0.001	0.001	0.001	0.001
Graduate	(0.001)	(0.001)	(0.001)	(0.001)
	0.006	0.005	0.004	0.003
Observations	13 538	25 524	44 696	54 101

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the overall sample. 2. CES-D score thresholds for 25th, 50th, 75th and 90th percentiles are 7, 9, 12 and 15 respectively. 3. Includes adults 45 and older. 4. Excluded if responses were reported by a proxy or had missing values for education variable. 5. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in past year and whether respondent visited a healthcare facility in past year. 6. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 7. Standard errors in parentheses. 8. Source: authors' construction based on LASI Wave I (weighted).

and behaviours by income levels (Vellakkal *et al.*, 2013; Onur and Velamuri, 2018). We therefore also estimate the models separately by income levels. We use the 25th percentile of the annual earnings distribution in our sample (only among those who have any income at all), which is INR 36,000 (about US$1.27 per day).

When we estimate the regressions separately by income threshold of INR 36,000 annual earnings, the main results hold true primarily for the group above the threshold (Table 10) but not necessarily for those below (Table 9). Among those who earn less than INR 36,000 a year, there are no significant differences in reporting error by education levels for accurate reporting of high blood pressure. However, those with college education and graduate degrees are significantly more likely to report that they do not have mental depression when the CES-D scale result shows otherwise.

It is possible that a person might be more aware of their health condition if their family members have that particular condition. This could lead some to get health checks more often. The data allow us to identify individuals who have any family members with high blood pressure. We thus estimate the main model with additional controls on income- and health-related variables (as in Tables 3 and 4) among those who report having any family members with high blood pressure. As shown in Table 11, the result on false negative reporting remains similar to our other results, i.e. the less educated are more likely to give a false negative report of high blood pressure. Those with no education were less likely to give accurate report of high blood pressure. However, the results for accurate reporting show a slightly weaker relationship by education levels than in our main findings.

Unlike asymptomatic conditions such as high blood pressure, whether one engages in physical activity is a conscious behaviour that does not require any diagnosis. We thus estimate the model for the group of respondents that accurately report their physical activity in the past month. There are several questions related

Table 7: Linear probability regression results for accurate, false negative and false positive reporting of health condition (male only) with income and health controls

Variables	Accurate reporting		False negative reporting		False positive reporting	
	High blood pressure	Mental depression	High blood pressure	Mental depression	High blood pressure	Mental depression
No education	−0.023*	−0.070***	0.161***	0.002	−0.016*	0.001
	(0.012)	(0.013)	(0.020)	(0.003)	(0.010)	(0.001)
Less than middle school	−0.028**	−0.054***	0.093***	0.004	0.006	0.002
	(0.012)	(0.012)	(0.018)	(0.004)	(0.011)	(0.002)
Bachelor’s	0.027	−0.027	−0.063**	−0.000	0.000	0.000
	(0.022)	(0.026)	(0.029)	(0.005)	(0.018)	(0.002)
Graduate	−0.086	0.128***	−0.024	−0.006	0.010	0.005
	(0.066)	(0.033)	(0.043)	(0.011)	(0.025)	(0.005)
Observations	26 672	28 611	11 683	11 549	14 989	17 180

Notes 1. This table displays linear probability regression results on gender-specific determinants of health condition and behaviour reporting. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors’ construction based on LASI Wave I (weighted).

Table 8: Linear probability regression results for accurate, false negative, and false positive reporting of health condition (female only) with income and health controls

Variables	Accurate reporting		False negative reporting		False positive reporting	
	High blood pressure	Mental depression	High blood pressure	Mental depression	High blood pressure	Mental depression
No education	−0.014	−0.033	0.077**	−0.002	0.010	0.000
	(0.024)	(0.024)	(0.030)	(0.004)	(0.011)	(0.002)
Less than middle school	−0.002	−0.002	0.003	0.001	0.024**	0.001
	(0.023)	(0.024)	(0.028)	(0.004)	(0.011)	(0.002)
Bachelor’s	0.036	0.050	−0.055	0.008*	0.034	0.001
	(0.041)	(0.048)	(0.066)	(0.005)	(0.031)	(0.003)
Graduate	−0.025	−0.061	0.013	−0.006	0.044	−0.003*
	(0.045)	(0.087)	(0.067)	(0.008)	(0.045)	(0.002)
Observations	30 935	33 071	14 122	15 128	16 813	18 047

Notes 1. This table displays linear probability regression results on gender-specific determinants of health condition and behaviour reporting. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors’ construction based on LASI Wave I (weighted).

Table 9: Linear probability regression results for accurate, false negative and false positive reporting of health condition (among those reporting earnings < INR 36,000/year) with work status and health controls

Variables	Accurate reporting		False negative reporting		False positive reporting	
	High blood pressure	Mental depression	High blood pressure	Mental depression	High blood pressure	Mental depression
No education	−0.014	−0.057***	0.106***	−0.000	−0.004	0.001
	(0.016)	(0.017)	(0.022)	(0.004)	(0.009)	(0.001)
Less than middle school	−0.004	−0.018	0.028	0.005	0.012	0.001
	(0.015)	(0.017)	(0.020)	(0.004)	(0.009)	(0.002)
Bachelor’s	0.017	0.028	−0.035	0.012**	0.027	−0.001
	(0.027)	(0.035)	(0.037)	(0.005)	(0.023)	(0.002)
Graduate	0.012	0.100***	−0.058	0.014*	−0.000	−0.001
	(0.030)	(0.031)	(0.043)	(0.008)	(0.031)	(0.003)
Observations	37 469	39 906	18 009	18 135	19 460	21 926

Notes 1. This table displays linear probability regression results on poverty status-specific determinants of health condition and behaviour reporting. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors’ construction based on LASI Wave I (weighted).

Table 10: Linear probability regression results for accurate, false negative, and false positive reporting of health condition (among those reporting earnings ≥ INR 36,000/year) with work status and health controls

Variables	Accurate reporting		False negative reporting		False positive reporting	
	High blood pressure	Mental depression	High blood pressure	Mental depression	High blood pressure	Mental depression
No education	−0.035**	−0.050***	0.165***	0.001	−0.003	0.000
	(0.016)	(0.017)	(0.027)	(0.002)	(0.012)	(0.001)
Less than middle school	−0.037**	−0.056***	0.121***	0.001	0.015	0.001
	(0.015)	(0.016)	(0.024)	(0.002)	(0.015)	(0.001)
Bachelor's	0.032	−0.042	−0.094**	−0.007	−0.004	0.002
	(0.029)	(0.034)	(0.039)	(0.006)	(0.019)	(0.002)
Graduate	−0.147	0.057	0.027	−0.014	0.027	0.008
	(0.091)	(0.068)	(0.061)	(0.010)	(0.031)	(0.006)
Observations	20 138	21 776	7796	8542	12 342	13 301

Notes 1. This table displays linear probability regression results on poverty status-specific determinants of health condition and behaviour reporting. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors' construction based on LASI Wave I (weighted).

Table 11: Linear probability regression results for accurate, false negative and false positive reporting of health condition(among those who have any family members with high blood pressure) with income and health controls

Variables	Accurate reporting High blood pressure	False negative reporting High blood pressure	False positive reporting High blood pressure
No education	−0.020*	0.108***	0.000
	(0.010)	(0.018)	(0.007)
Less than middle school	−0.017	0.043**	0.019**
	(0.010)	(0.017)	(0.008)
Bachelor's	0.005	−0.054	0.028
	(0.022)	(0.033)	(0.018)
Graduate	−0.003	−0.095	0.013
	(0.031)	(0.058)	(0.022)
Observations	44 170	18 608	25 562

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the sample of individuals with a family history of high blood pressure. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors' construction based on LASI Wave I (weighted).

to physical activity in the data, such as whether the respondent has engaged in vigorous or moderate physical activity in the past month and whether they walked the previous day. We use the question on how often they engaged in sports or vigorous physical activities in the past month. We examine different variables for self-reports of physical activity in the analysis but mainly focus on whether the respondent reports having engaged in physical activity in the past month as a general proxy for physical activity.

For objective measurement on physical activity, we use the data on timed walking, measured among older adults who do not have a health condition that limits walking ability. Respondents walk along a 4m path in a non-carpeted area (with a measuring tape indicating the walking space) at their usual pace of walking. The measurement is undertaken twice, with an interviewer present. We calculate the average time of walking measured in seconds and divide it by four to calculate the walking speed (seconds per metre). We proxy for objective measurement of physical activity (specifically walking) based on the walking speed from the timed walk (1 m/s or greater) (Middleton *et al.*, 2015). The threshold of 1 m/s was chosen based on associated outcomes such as being independent and less likely to be hospitalized or have adverse events among older adults (Middleton *et al.*, 2015).

Table 12 shows the regression results for accurate, false negative and false positive reporting of health condition among those who accurately report their physical activity engagement in the past month. This sample includes older adults who report not engaging in physical activity in the past month and who do not have a walking speed above the threshold, and the opposite case (those who exercised in the past month and had a walking speed above the threshold). It is assumed that this will be the subgroup that correctly recalls their exercise level and accurately reports it in the data. Specifically, accurate reporting of high blood pressure does not seem to differ by education levels. For outcomes on depression, the main findings remain the same: those with no schooling or less than middle school education are less likely to accurately report than those with middle or high school education. False negative reporting of depression is significant and the coefficient for false negative reporting is largest for those with college education. However, false negative reporting did not significantly differ by education levels among the group that did not accurately report physical activity. This might indicate that among those respondents who accurately recall their recent exercise level and are more likely to accurately report conscious behaviour, social desirability

Table 12: Linear probability regression results for accurate, false negative and false positive reporting of health condition (individuals with accurate report of physical activity in the past month) with income and health controls

	Accurate reporting		False negative reporting		False positive reporting	
Variables	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**	**High blood pressure**	**Mental depression**
No education	−0.019	−0.067***	0.105***	−0.000	0.000	0.001
	(0.018)	(0.018)	(0.024)	(0.003)	(0.007)	(0.001)
Less than middle school	−0.004	−0.034**	0.030	0.009**	0.019**	0.002**
	(0.016)	(0.017)	(0.021)	(0.004)	(0.008)	(0.001)
Bachelor’s	0.013	−0.006	−0.050	0.012**	0.028	0.001
	(0.029)	(0.031)	(0.039)	(0.005)	(0.018)	(0.002)
Graduate	0.006	0.052	−0.079*	0.010	0.013	0.008
	(0.030)	(0.032)	(0.043)	(0.007)	(0.022)	(0.007)
Observations	44 170	47 252	18 608	20 424	25 562	26 994

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting for the subsample of individuals who accurately report their levels of physical activities in the last month. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, state fixed effects, annual earnings in rupees, current work status, whether respondent has insurance coverage, hospitalization in the past year and whether respondent visited a healthcare facility in past year. 5. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 6. Standard errors in parentheses. 7. Source: authors’ construction based on LASI Wave I (weighted).

Table 13: Linear probability regression results for accurate reporting of health condition among those who have been diagnosed with hypertension or depression within the last 5 years

	Accurate reporting	
Variables	**High blood pressure**	**Mental depression**
No education	−0.067***	−0.029
	(0.019)	(0.077)
Less than middle school	−0.042**	−0.171**
	(0.020)	(0.077)
Bachelor’s	0.022	−0.022
	(0.037)	(0.225)
Graduate	0.032	-0.153
	(0.041)	(0.165)
Observations	9556	709

Notes 1. This table displays linear probability regression results on the determinants of health condition and behaviour reporting among those who have been diagnosed with hypertension or depression within the last 5 years. 2. Includes adults 45 and older. 3. Excluded if responses were reported by a proxy or had missing values for education variable. 4. Regressions control for number of words recalled correctly as well as indicator variables of age, gender, marital status, religion, first language, whether respondent lives in a rural area, and state fixed effects, work status and annual earnings in rupees. 5. The truncated sample size is too small to provide the results for false positive and false negative reports of high blood pressure and depression among those who are diagnosed within the last 5 years.6. ***$P < 0.01$, **$P < 0.05$, *$P < 0.10$. 7. Standard errors in parentheses. 8. Source: authors’ construction based on LASI Wave I (weighted).

bias might be driving the results on false negative reporting of depression.

Finally, a concern might be that the diagnosis of either high blood pressure or depression was done a long time back, and the respondent might well be cured of the disease since that time. If this is indeed the case then the estimates of false-positive reporting in particular (where the respondent answers ‘yes’ to whether they have a disease while medical tests show otherwise) may exhibit an upward bias. To address this issue, we truncate the sample to only those who have been diagnosed with high blood pressure or depression within the past 5 years. Tables 13 reports the results, and we find that our results are qualitatively similar to that in Table 2 for the overall sample. Respondents with no education are less likely to report either disease accurately.

Conclusion

This is the first study to analyze the health–education gradient for older adults in India based on differential health reporting error. This is also the first study to analyze the accuracy of self-diagnosis for depression in a developing country. We find evidence suggesting that older adults with no schooling and less than middle school education report high blood pressure and depression less accurately than those with middle or high school education. This pattern is consistent across different controls, such as earnings, healthcare utilization and community-based identifiers such as scheduled caste status. We find false negative reporting for high blood pressure to be dominant (where someone over-reports their health, e.g. stating an absence of high blood pressure in the self-report but where the objective measurement indicates otherwise) for those with no education. Differential levels of inaccuracy in self-health assessment depending on education status, and false negative assessment by those with low levels of education specifically, have important public policy implications. In particular, public investment in improving health outcomes should be sensitive to potential under-investment in health by select segments of the population due to their erroneous self-health assessments. Education campaigns and publicly funded health check-ups catering to populations particularly vulnerable to either ignoring symptoms and/or optimistic self-assessments of health status should be a priority in a developing country such as India.[12]

Sensitivity analysis for depression shows our baseline results based on the 10-item CES-D questionnaire to be robust, with the less educated less likely to accurately report depression at all

[12] Several possibilities might be at play behind false negative and false positive reporting. First, the diagnosis might have been many years ago and the respondent could well be cured of the disease at the time of the interview for false positive, and vice versa for false negative (diagnosed not having the disease a long time back and have since developed it). Second, the earlier diagnosis itself might have been faulty caused by defective instruments or unique circumstances faced by the respondent at the time of the earlier or current diagnosis (job loss, stress, other underlying own-health and family issues, etc.) that can affect the incidence of both false positive and false negative reporting. Third, for false negative reporting, there might be a behavioural interpretation as in Sen (1999): ‘*Our desires and pleasure-taking abilities adjust to circumstances…those who are persistently deprived… the routinely overworked sweatshop worker in exploitative conditions…tend to come to terms with their deprivation.*’ Thus, an individual may very well come to terms/normalize a medical condition depending upon their environment and circumstances.

thresholds than those who have middle and high school education. Furthermore, our results show that men with no schooling and less than middle school education report high blood pressure and depression less accurately: women with no schooling are significantly more likely to make false negative reports regarding high blood pressure than those with middle or high school education. Surprisingly, our finding that older adults with no schooling and less than middle school education are more likely to be less accurate in their reporting for high blood pressure and depression holds true only for those whose income is above the poverty line of roughly US $1.27 per day.

SUPPLEMENTARY DATA

Supplementary data are available at *Oxford Open Economics* online.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

This paper has benefited from comments by seminar participants at UNU-WIDER, Helsinki, and session participants at the Southern Economics Association Conference and European Health Economics Association Conference. We are particularly grateful to Omar McDoom, Rachel Gisselquist, Rahul Lohati, Kunal Sen, Enrico Nichelatti, Rodrigo Oliveira and two anonymous referees for their insightful suggestions. All remaining errors and omissions are ours.

REFERENCES

Andresen, E. M. *et al.* (1994) 'Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D', *American Journal of Preventive Medicine*, **10**: 77–84.

Bernadine, M. (2021) 'Understanding India's Mental Health Crisis', *Ideas for India*, 6 April, Accessed 27 Jan 2023.

Bhan, N. *et al.* (2020) 'Access to Women Physicians and Uptake of Reproductive, Maternal and Child Health Services in India', *EClinical Medicine*, **20**: 100309.

Birrer, R. B., and Vemuri, S. P. (2004) 'Depression in Later Life: a Diagnostic and Therapeutic Challenge', *American Family Physician*, **69**: 2375–82.

Bloom, D., Sekher, T. V., and Lee, J. (2021) 'Longitudinal Aging Study in India (LASI): New Data Resources for Addressing Aging in India', *Nature Aging*, **1**: 1070–2.

Carey, M. *et al.* (2015) 'Unassisted Detection of Depression by GPs: Who Is most Likely to be Misclassified?' *Family Practice*, **32**: 282–7.

Choi, A., and Cawley, J. (2018) 'Health Disparities across Education: The Role of Differential Reporting Error', *Health Economics*, **27**: e1–29.

Coelho, K. *et al.* (2016) 'A Systematic Review of the Literature on the Relationship between Caste Membership and Health-Related Risk Factors in India', *International Journal of Medicine and Public Health*, **6**: 61–8.

Cramm, J. M. *et al.* (2015) 'The Health of india's Elderly Population: a Comparative Assessment Using Subjective and Objective Health Outcomes', *Journal of Population Ageing*, **8**: 245–59.

Cutler, D. M., and Lleras-Muney, A. (2010) 'Understanding Differences in Health Behaviors by Education', *Journal of Health Economics*, **29**: 1–28.

Dandona, L. *et al.* (2020) 'The Burden of Mental Disorders across the States of India: the Global Burden of Disease Study 1990–2017', *Lancet Psychiatry*, **7**: 148–61.

Devita, M. *et al.* (2022) 'Recognizing Depression in the Elderly: Practical Guidance and Challenges for Clinical Management', *Neuropsychiatric Disease and Treatment*, **Volume 18**: 2867–80.

Dey, S. *et al.* (2012) 'Health of the elderly in India: Challenges of access and affordability', in Smith, J. P., and Majmundar, M., (eds.) *Aging in Asia: Findings from New and Emerging Data Initiatives*, Washington, DC: National Academies Press.

Geldsetzer, P. *et al.* (2018) 'Diabetes and Hypertension in India: a Nationally Representative Study of 1.3 Million Adults', *JAMA Internal Medicine*, **178**: 363–72.

Gellis, Z. D. (2010) 'Assessment of a Brief CES-D Measure for Depression in Homebound Medically Ill Older Adults', *Journal of Gerontological Social Work*, **53**: 289–303.

Grossman, M. (2008) 'The Relationship between Health and Schooling on JSTOR', *Eastern Economic Journal*, **34**: 281–92.

Huffman, M. D. *et al.* (2011) 'Incidence of Cardiovascular Risk Factors in an Indian Urban Cohort: Results from the New Delhi Birth Cohort', *Journal of the American College of Cardiology*, **57**: 1765–74.

IIPS (2023) '*Longitudinal Ageing Study in India (Lasi): Datasets*. Available at: www.iipsindia.ac.in/content/LASI-data accessed 27 January 2023.

IIPS, National Programme for Health Care of Elderly 'Indian Ministry of Health and Family Welfare, Harvard T.H. Chan School of Public Health and University of Southern California (2020)', in *Longitudinal Aging Study in India (LASI) Wave 1, 2017–18, Executive Summary*Mumbai.

Iqbal, S. *et al.* (2022) 'Smoking Cessation Interventions in South Asian Region: A Systematic Scoping Review', *BMC Public Health*, **22**: 1096.

Johnston, D. W., Propper, C., and Shields, M. A. (2009) 'Comparing Subjective and Objective Measures of Health: Evidence from Hypertension for the Income/Health Gradient', *Journal of Health Economics*, **28**: 540–52.

Krishnamoorthy, Y., Rajaa, S., and Rehman, T. (2020) 'Diagnostic Accuracy of Various Forms of Geriatric Depression Scale for Screening of Depression among Older Adults: Systematic Review and Meta-Analysis', *Archives of Gerontology and Geriatrics*, **87**: 104002. https://doi.org/10.1016/j.archger.2019.104002

Lee, J. *et al.* (2015) 'Education, Gender, and State-Level Disparities in the Health of Older Indians: Evidence from Biomarker Data', *Economics & Human Biology*, **19**: 145–56.

Lowes, S., and Montero, E. (2021) 'The Legacy of Colonial Medicine in Central Africa', *American Economic Review*, **111**: 1284–314.

Middleton, A., Fritz, S. L., and Lusardi, M. (2015) 'Walking Speed: the Functional Vital Sign', *Journal of Aging and Physical Activity*, **23**: 314–22.

Najnin, N. *et al.* (2019) 'Impact of a Large-Scale Handwashing in- Tervention on Reported Respiratory Illness: Findings from a Cluster-Randomized Controlled Trial', *American Journal of Tropical Medicine and Hygiene*, **100**: 742–9.

Onur, I., and Velamuri, M. (2018) 'The Gap between Self-Reported and Objective Measures of Disease Status in India', *PLoS One*, **13**: e0202786.

Pednekar, M. S., Gupta, R., and Gupta, P. C. (2011) 'Illiteracy, Low Educational Status, and Cardio- Vascular Mortality in India', *BMC Public Health*, **11**: 567.

Perianayagam, A. *et al.* (2022) 'Cohort Profile: The Longitudinal Aging Study in India (LASI)', *International Journal of Epidemiology*, **51**: e167–76.

Radloff, L. S. (1977) ‘The CES-D Scale: A Self-Report Depression Scale for Research in the General Population’, *Applied Psychological Measurement*, **1**: 385–401.

Raghupathi, W., and Raghupathi, V. (2020) ‘The Influence of Education on Health: An Empirical Assessment of OECD Countries for the Period 1995–2015’, *Archives of Public Health*, **78**: 20.

Sen, A. (1999) *Development as Freedom*, New York: Knopf.

Smith, R., and Meeks, S. (2019) ‘Screening Older Adults for Depression: Barriers across Clinical Discipline Training’, *Innovation in Aging*, **3**:1–10

Van der Heide, I. *et al.* (2013) ‘The Relationship between Health, Education, and Health Literacy: Results from the Dutch Adult Literacy and Life Skills Survey’, *Journal of Health Communication*, **18**: 172–84.

Vellakkal, S. *et al.* (2013) ‘Socioeconomic Inequalities in Non-communicable Diseases Prevalence in India: Disparities between Self-Reported Diagnoses and Standardized Measures’, *PLoS One*, **8**: e68219.

World Bank (2019) *World Health Organization Global Health Expenditure Database*. https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS. Accessed 30 August 2023.