PROGRAM ON THE GLOBAL DEMOGRAPHY OF AGING AT HARVARD UNIVERSITY

Working Paper Series

Education, Gender, and State-Level Gradients in the Health of Older Indians: Evidence from Biomarker Data

Jinkook Lee, Mark E. McGovern, David E. Bloom, P. Arokiasamy, Arun Risbud, Jennifer O'Brien, Varsha Kale, and Peifeng Hu

February 2015

Education, Gender, and State-Level Gradients in the Health of Older Indians: Evidence from Biomarker Data[*]

Jinkook Lee[†1,2], Mark E. McGovern[3,4], David E. Bloom[4], P. Arokiasamy[5], Arun Risbud[6], Jennifer O'Brien[4], Varsha Kale[6], and Peifeng Hu[7]

[1]Dornsife Center for Economic and Social Research, University of Southern California, Los Angeles, USA
[2]RAND Corporation, Santa Monica, USA
[3]Harvard Center for Population and Development Studies, Cambridge, USA
[4]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, USA
[5]International Institute for Population Sciences, Mumbai, India
[6]National AIDS Research Institute, Maharashtra, India
[7]University of California, Los Angeles, USA

February 2015

Abstract

This paper examines health disparities in biomarkers among a representative sample of Indians aged 45 and older, using data from the pilot round of the Longitudinal Aging Study in India (LASI). Hemoglobin level, a marker for anemia, is lower for respondents with no schooling (0.7 g/dL less in the adjusted model) compared to those with some formal education. There are also substantial state and education gradients in underweight and overweight. The oldest old have higher levels of C-reactive protein (CRP) (1.1 mg/L greater than those aged 45-54), an indicator of inflammation and a risk factor for cardiovascular disease, as do those with greater body-mass index (an additional 1.2 mg/L for those who are obese compared to those who are of normal weight). We find no evidence of educational or gender differences in CRP, but respondents living in rural areas have CRP levels that are 0.8 mg/L lower than urban areas. We also find state-level disparities, with Kerala residents exhibiting the lowest CRP levels (1.96 mg/L compared to 3.28 mg/L in Rajasthan, the state with the highest CRP). We use the Blinder-Oaxaca decomposition approach to explain group-level differences, and find that state-level gradients in CRP are mainly due to heterogeneity in the association of the observed characteristics of respondents with CRP, as opposed to differences in the distribution of endowments across the sampled state populations.

JEL Classification: I12, I14, D30, O15

Keywords: Biomarkers, Health Disparities, Gender Differences, Blinder-Oaxaca Decomposition, Aging

1 Introduction

Many developing countries are currently undergoing rapid demographic and economic transitions. In particular, the proportion of individuals over the age of 50 in lower- and middle-income nations is expected to rise rapidly in coming decades (Shetty, 2012). This is the result of an ongoing epidemiological transition in these countries in which life expectancy is rising and mortality is shifting more towards later life (Prentice, 2006). The main risks for premature death are no longer solely the well-studied problems associated with poverty, such as malnutrition and poor sanitation. In India and China non-communicable diseases (NCDs), such as cardiovascular disease (CVD), which had been largely limited to higher-income countries, are becoming the main causes of premature mortality (Kearney et al., 2005). For example, NCDs now account for 60% of all deaths in India (World Health Organization, 2014), and the contribution to mortality of non-communicable relative to communicable diseases worldwide is expected to rise substantially in coming years. Given that India and China comprise one third of the world's population, how these changes alter well-being in these two countries will greatly affect global welfare. As well as being inherently important health outcomes, these diseases have important economic consequences. Non-communicable diseases exert large monetary and non-monetary impacts on society (Bloom et al., 2014a). NCDs also make a substantial contribution to morbidity, accounting for 54% of healthy life years (DALYs) lost in India (Murray et al., 2013). At the same time, malnutrition and communicable diseases remain substantial health threats in lower- and middle-income countries (Narayan et al., 2010). For example, 70% of women and children in India currently suffer from anemia (Balarajan et al., 2011a).

Rapid economic growth has contributed to improvements in living standards, but has also led to changing patterns of urbanization, diet, and other modifiable risk behaviors (Prentice, 2006). Coupled with existing excess mortality due to communicable disease, these differences in demographic, behavioral, and economic circumstances create a challenging environment for public health officials in India and other developing countries (Balarajan et al., 2011b). This challenge is exacerbated by difficulties in measuring the extent of health inequality and identifying the most at risk populations. There are major differences in both health and access to health care by gender, region, and level of education and other socioeconomic measures within countries. Understanding group-level differences in health outcomes is therefore important for establishing policy priorities.

Given that NCDs are increasingly affecting a growing proportion of the global population (Lopez et al.,

2006), their impact is attracting greater attention. Cardiovascular disease is now the second-most important contributor to mortality in India, accounting for 28% of deaths (World Health Organization, 2005). Yet, CVD has been a controversial subject in public discourse due to the perception that heart disease is mainly a problem of the urban upper middle class. The academic literature has debated the strength of the evidence for this claim (Gwatkin, 2013; Lloyd-Sherlock et al., 2014; Subramanian et al., 2013). While it has been argued that CVD may increasingly affect all socioeconomic groups, the problem of inadequate nutrition certainly remains a problem for a substantial proportion of the less well-off in India, especially for women (Bentley and Griffiths, 2003). India has the highest incidence of anemia in the world, with levels that have remained static for the past decade, despite economic growth (Balarajan et al., 2011a). Both cardiovascular disease and malnutrition thus have a substantial impact on public health in India, and also in other lower- and middle-income countries too. However, good evidence on the prevalence of malnutrition and CVD in India can be difficult to obtain, particularly for older Indians, and particularly at the regional level, where there is varying access to medical services and diagnosis. There are two reasons for this.

First, self-reporting on health in India is particularly problematic due to differential state-level access to care, diagnosis, and treatment. This access affects the extent to which individuals are aware of their health status, their recall bias, and differences in how they perceive their health (Johnston et al., 2009). This can result in heterogeneity in the thresholds used by respondents for indicating that they suffer from a medical condition (Sen, 2002). For example, there are large differences in hypertension diagnosis evident in self-reported and measured hypertension, and these discordances vary by state (Lee et al., 2012). Figure 1 illustrates these disparities by educational level with data from the Longitudinal Aging Study in India (LASI) pilot. Using self-reported measures to assess socioeconomic gradients in disease prevalence can be especially misleading (Vellakkal et al., 2013).

Second, among lower- and middle-income countries, there is a dearth of nationally representative data on the objective biological markers of malnutrition and CVD. In particular, there is little existing evidence from these countries about the risks for NCDs among persons aged 45 years and older, the population most likely to be affected by these conditions (Chaves et al., 2005). Even if health service records are available, they may not provide a complete picture of population health because they only provide information on people seeking diagnosis or treatment. Such individuals may not be a representative sample of the population, especially if access to health care is low. Although much analysis focuses on countries as a whole, regional differences within them are likely to be equally important. For example, India is a union of 29 states and 6 territories,

3

Figure 1: Educational Disparities in Measured and Reported Hypertension in India
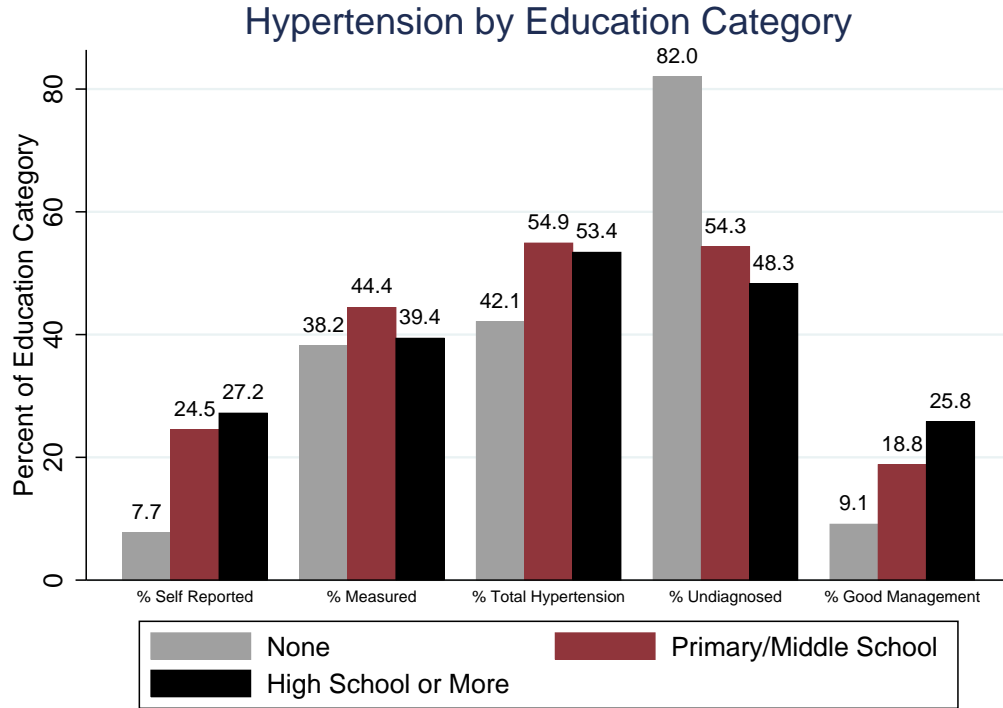
Note for Figure 1: The percent of each education category (no education, primary/middle school, high school or more) with self-reported hypertension, measured hypertension (blood pressure readings are higher than the following thresholds: systolic $\geq$ 140 mm Hg or diastolic $\geq$ 90 mm Hg), total hypertension (either self-reported or measured hypertension), undiagnosed hypertension (not diagnosed according to self-reports but measured hypertension), and good management of their hypertension (diagnosed to be hypertensive but blood pressure readings are lower than the above thresholds). Data are from the LASI pilot. Source: Lee et al. (2012), National Academy of Science.

which vary greatly in their economic development, cultures, education levels, and policies (Deaton and Dreze, 2002; Deaton and Kozel, 2005; Ravallion and Datt, 2002; Lee and Smith, 2014). These differences may lead to cross-state variation in risks for these health outcomes.

Assessing health outcomes accurately is important for understanding the successes and failures of alternative policies and environments, as well as for understanding where to target interventions. Differential health and economic outcomes are evident when comparing urban and rural groups, states, and genders. Table 1 illustrates some of these disparities in the four Indian states covered by the LASI pilot. While all four of these states saw urban consumption increase substantially between the 1960s and the 1990s, there was little improvement in rural consumption between 1960-1961 and 1993-1994, with the exception of Kerala. Kerala also had more favorable statistics for females, including a higher ratio of girls to boys and higher school attendance rates.

Using biomarkers to directly assess risks for particular outcomes provides a potential solution to the lack of good health information, while also providing an immediate assessment of objective health disparities for both individuals and groups. Given the increasing importance of heart disease and the well-documented lack of adequate nutrition (Bentley and Griffiths, 2003), and as we outline above, CVD and anemia are especially important public health issues in India. Moreover, the impact of these conditions is likely to be the most harmful among older individuals (Carmel, 2001; Chaves et al., 2005).

Two attractive candidates for targeted biomarker data collection are C-reactive protein (CRP) and hemoglobin (Hb). CRP is a biomarker for inflammation, which is associated with increased risk of cardiovascular disease (Vikram et al., 2003). Elevated levels of CRP are also associated with hypertension and diabetes, with thresholds for high risk defined by the Center for Disease Prevention and Control (CDC) and the American Heart Association (AHA) (Myers et al., 2004). The Hb biomarker can be used to evaluate the prevalence of anemia (Balarajan et al., 2011a). Anemia results from a lack of either red blood cells or hemoglobin, and leads to weakness or fatigue (Aguayo et al., 2003; Beghé et al., 2004; Denny et al., 2006). In developing countries, anemia is often associated with iron or vitamin deficiencies due to poor nutrition.

Data collection on the prevalence of anemia has mainly focused on preschool-age children, pregnant women, and non-pregnant women of reproductive age, particularly in India (Balarajan et al., 2013; Bentley and Griffiths, 2003; Ghosh, 2009). The World Health Organization (WHO) does not report country-level estimates for school-age children, men, and the elderly for this reason (de Benoist et al., 2008). Initiatives such as LASI, which recently collected information on a variety of biomarkers from the elderly, are helping shift the focus of data collection. LASI is designed to be representative of both India as a whole and of its constituent states, and will ultimately follow more than 50,000 respondents longitudinally (Arokiasamy et al., 2012). The survey includes respondents aged 45 and older, as well as their spouses (regardless of age).

In 2010, LASI collected pilot data in four states: Punjab, Rajasthan, Kerala, and Karnataka, interviewing 1,683 eligible individuals. Of these, 78% (1,305 respondents) provided a dried blood spot (DBS) sample. Individual and household micro data from the survey are publicly available, and the biomarker data are also accessible through an application for a restricted data file (at the University of Southern California Gateway to Global Aging Data website: www.g2aging.org). One of LASI's main contributions is to assess health status using objective biomarker data from nationally and in-state representative samples of older Indians. In this paper, we use the 2010 pilot data to address the following research questions.

Table 1: Changes in Economic and Social Factors by State

	Punjab	Rajasthan	Karnataka	Kerala
Economic Growth				
1960-61 Mean Per Capita Consumption (Rs/Month)[1]				
Rural	82.06	55.7	59.19	46.64
Urban	83.71	66.6	72.05	53.6
1993-94 Mean Per Capita Consumption (Rs/Month)[1]				
Rural	79.23	58.07	62.52	73.44
Urban	100.34	75.87	79.82	89.32
Male Preference				
2001 Child sex ratio (Girls per 1,000 boys aged 0 - 6)[2]	798	909	946	960
2008 Elementary School Attendance Rates (Ages 5 - 14) Per 1,000[3]				
Boys	897	847	898	968
Girls	882	710	866	985

Source:
1. (Datt, 1998)
2. The Government of India 2011 Census
3. The Government of India, Ministry of Human Resource Development, Selected Educational Statistics: 2000-01

First, we establish whether there is an education gradient for hemoglobin. Given the existing literature, we hypothesize the existence of socioeconomic disparities for this outcome (Bentley and Griffiths, 2003). Second, we examine the risk factors associated with anemia separately for men and women, given the much higher incidence of anemia among women (Balarajan et al., 2011a). Third, given the growing prevalence of obesity in developing countries and its association with NCDs (Ackerson et al., 2008; Kearney et al., 2005; Lee et al., 2012; Popkin et al., 2001; Prentice, 2006), we also examine the risk factors for being underweight and overweight. We hypothesize similar education gradients in BMI to the gradients we observe for hemoglobin (Subramanian et al., 2009). Fourth, given the public discourse on socioeconomic status (SES) and cardiovascular disease, we assess whether there is an education gradient in CRP. We also examine differences between urban and rural areas. Although the relationship between SES and some biomarkers may be context specific (Beltrán-Sánchez and Crimmins, 2013; Crimmins, 2015), previous research found no evidence of an SES gradient in CRP in Costa Rica (Rosero-Bixby and Dow, 2009), another country experiencing demographic, economic, and epidemiologic transitions. Therefore, we hypothesize that we will

find no SES gradient in CRP in India. Finally, given the varying growth and social policies of Indian states in recent decades, we examine state-level variation in CRP. We aim to determine whether differences in economic growth and social policies contribute to state-level variation in CRP.

The rest of this paper is structured as follows. In Section 2, we discuss our data and analytic approach. In Section 3, we present our results. We conclude in Section 4.

2 Data and Analytic Approach

The 2010 LASI pilot sample was drawn using a stratified, multistage, area probability sampling design based on the 2001 Indian Census. From each state, we randomly chose two Census 2001 districts. We then randomly selected eight primary sampling units (PSU) from each district to match the urban/rural share of the state population. Finally, we selected 25 community-residing households through random sampling from each PSU. For external validity, it is important that the LASI pilot data provide a reasonable approximation to the population of interest. Arokiasamy et al. (2012) conducted an analysis of whether the LASI pilot data derived from two districts in each state was comparable to the older population in India as a whole, and the states which were surveyed in LASI. Overall, the characteristics of older respondents in LASI (aged 45 and over) closely matched the characteristics of older respondents in the National Sample Survey (NSS), the India Human Development Survey (IHDS), the World Health Survey (WHS), and the WHO Study on global AGEing and adult health (SAGE). SAGE was conducted in two of the states covered by LASI, Rajasthan and Karnataka (Kowal et al., 2012), permitting a direct comparison of respondents in both surveys in these states. LASI was also found to provide a good match with the population of interest on the basis of this metric, for example, the age structure in LASI for Rajasthan was 43.1% (45-54), 23.4% (55-64), 21.8% (65-74) and 11.8% (75+), while the age structure in SAGE for Rajasthan was 49.9% (45-54), 26.9% (55-64), 16.3% (65-74), and 6.8% (75+). Further details of this analysis are provided in Arokiasamy et al. (2012), available at www.ncbi.nlm.nih.gov/books/NBK109220/.

Survey weights are provided with the LASI pilot data, based on the 2011 Indian Census. One set of weights matches the biomarker sample with the population aged 45 and older in the four surveyed states (Punjab, Rajasthan, Kerala, and Karnataka) based on age, sex, and urban/rural place of residence. A second set of weights matches the biomarker sample to the population aged 45 and older in India as a whole. We use these latter weights in the analysis, however we have verified that results are very similar when using state

weights or when not weighting the data.

LASI has two main modules: the household and individual interview, and the biomarker collection. The household interview asks about physical environment and household finances. The individual interview asks about demographics, family, social activities, health and health behaviors, and work and pensions. For the collection of dried blood spots (DBS), respondents provided separate consent, permitting interviewers to prick their finger and place five drops of blood on a Whatman 903 Protein Saver card. The collected DBS cards were left to air dry for at least 4 hours for overnight, then sent to the National AIDS Research Institute (NARI) in Pune, India, where they were stored below -20° C and later assayed.

Both CRP concentration and hemoglobin levels in the DBS specimens were measured using validated methods. CRP concentration was measured using the enzyme-linked immunosorbent assay (ELISA) protocol developed by McDade et al. (2004), and hemoglobin levels were measured using the method developed by O'Broin and Gunter (1999). To ensure quality, all samples, standards, and controls were measured in duplicate. Internal quality controls were run on every plate, and plates with out-of-range quality-control values were re-run. We further ensured the quality of laboratory assay results through periodic use of external quality control samples prepared by the USC/UCLA Center on Biodemography and Population Health.

We also externally validated work at the NARI laboratory. For CRP assays, we compared NARI's results on 32 validation samples with DBS-based values from the reference laboratory in the United States (at the University of Washington). The correlation coefficient was 0.95. For hemoglobin levels, 33 validation samples had DBS values from NARI and venous-based results from the UCLA Clinical Laboratory. The Pearson's correlation coefficient was 0.78. In general, NARI had higher Hb values than the corresponding venous-based results. The average difference was 0.55 g/dL (standard deviation: 0.86 g/dL). For measures of both Hb and CRP, DBS specimens were run in duplicate. The two values from duplicate measurements were very highly correlated. In the analysis in this paper, we combined the duplicate measures into an average. Using either the first or second measure on its own had little effect on the results. For further details of the LASI biomarker data collection and the external validation work done at the NARI laboratory, see Bloom et al. (2014b), available at www.rand.org/pubs/working_papers/WR1043.html.

Table 2 presents descriptive statistics for the analysis of the two main dependent variables (hemoglobin, measured in grams per deciliter, g/dL, and CRP, measured in milligrams per liter, mg/L) and independent variables (age, gender, state, caste, urban/rural residency, education, smoking, body mass index group, household living arrangements, and self-reported diagnosis of diabetes, stroke, hypertension, or a heart

condition). We focus on respondents over the age of 45, excluding a small number of spouses under this age. This left 1,150 observations in total. There were a small number of missing values for some covariates and outcomes, but these did not exceed 6% of observations. For example of the 1,150 total respondents, 1,077 had information on Hb. The mean Hb in the sample was 14.3 g/dL, and mean CRP was 2.7 mg/L. Table 2 also shows the proportion of respondents with anemia (20%) as determined by Hb levels (below 12 g/dL for women and below 13 g/dL for men), as well as those at high risk for cardiovascular disease (30%), as determined by CRP levels of more than 3 mg/L (Myers et al., 2004).

Nearly half the sample (46%) was between 45-54 years of age, nearly half (48%) were male, and more than half (55%) had received some formal schooling. There were roughly equal numbers of respondents in each state (ranging from 253 in Karnataka to 329 in Kerala). For covariates, we focus on pre-determined variables which are unlikely to be outcomes of socioeconomic status, health status, or health care use. Nevertheless, we also consider smoking, body mass index (BMI), and the presence of a self-reported chronic health conditions, as these have been associated with Hb and CRP (Beghé et al., 2004; Carmel, 2001; Daly, 2013; Danesh et al., 2004). Factors such as height and weight could be viewed as being consequences of health conditions or socioeconomic status in early life (Coffey, 2014; Fernihough and McGovern, 2015), however in our analysis we find that their addition to the model has little effect on our estimates. Living arrangements and marital status of older respondents have been shown to be associated with health and poverty (Bongaarts and Zimmer, 2002; Dreze and Srinivasan, 1997; Edmonds et al., 2005; Hu and Goldman, 1990; Hughes and Waite, 2002; Rahman, 2013). With the growing proportion of this demographic in countries such as India, identifying the ideal living situation for older individuals will assume a growing importance (Rajan and Kumar, 2003). Therefore, in the analysis we include a series of indicators for whether the respondent was married and living with a child, married and not living with a child (or children), not married and living with a child, and not married and not living with a child. Although the data differentiate between widowhood, divorce, separation and never having married, in the analysis we define the categories for marital status in terms of being currently married or not, as there are relatively few individuals who are divorced, separated or never married.

Figure 2 presents the distribution of CRP in the LASI pilot sample, showing the high risk cut-off at 3 mg/L. We show the combined sample, as we find no major gender differences for this outcome. The distribution is skewed to the right, although a substantial proportion has CRP above 3 mg/L (30%). Figures 3 and 4 show the Hb distribution stratified by gender, which in both cases approximates a normal distribution. The

distributions for BMI among men and women are shown in figures 5 and 6. 19% of men in the sample are overweight (BMI $\geq$ 25 kg/m^2), compared to 31% of women, whereas 26% of men are underweight (BMI <18.5 kg/m^2) compared to 19% of women.

Table 2: Descriptive Statistics for the Analysis Sample

	Median	Mean	SD	N
Hemoglobin (g/dL)	14.31	14.26	2.50	1,077
C-Reactive Protein (mg/L)	1.68	2.68	3.02	1,106
Body Mass Index (kg/m^2)	21.91	22.35	5.07	1,133

Age Group	No.	%	Residency	No.	%
44-54	532	46.3	Urban	298	25.9
55-64	318	27.7	Rural	852	74.1
65-74	202	17.6	Total	1,150	100
75+	97	8.4			
Total	1,149	100	**State**		
			Punjab	288	25
Gender			Rajasthan	280	24.3
Male	552	48	Kerala	329	28.6
Female	598	52	Karnataka	253	22
Total	1,150	100	Total	1,150	100
Caste			**Diabetes, Stroke, Hypertension or Heart Condition**		
Scheduled Caste	188	16.8	No	862	75
Scheduled Tribe	138	12.3	Yes	288	25
Other Backward Class	395	35.2	Total	1,150	100
None	400	35.7			
Total	1,121	100	**Smokes Now**		
			No	971	84.7
Education			Yes	176	15.3
Some Schooling	627	54.5	Total	1,147	100
No Schooling	523	45.5			
Total	1,150	100	**BMI Group**		
			Underweight (<18.5)	253	22.3
Household Living Arrangements			Normal (18.5 - 24.9)	592	52.3
Married, No Children in Household	160	13.9	Overweight (25 - 29.9)	218	19.2
Married, Children in Household	753	65.5	Obese (30+)	70	6.2
Not Married, Children in Household	61	5.3	Total	1,133	100
Not Married, No Children in Household	176	15.3			
Total	1,150	100			

Source: LASI Pilot 2010 biomarker sample. Those under age 45 are excluded. Underweight is an indicator for BMI<18.5, normal weight is 18.5 $\leq$ BMI <25, overweight is 25 $\leq$ BMI<30, and obese is BMI$\geq$ 30. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease. The sample is weighted.

We then examine population-level differences in hemoglobin and CRP, by estimating bivariate associations between these biomarkers and the explanatory variables shown in table 2. Tables 3 and 4 show this analysis for Hb and CRP, respectively. Each table shows, for each subpopulation, the mean level along with the associated standard error, the confidence interval, and a t-test for the difference between the mean level in that group and the relevant comparison category.

Figure 2: C-Reactive Protein Distribution for Men and Women
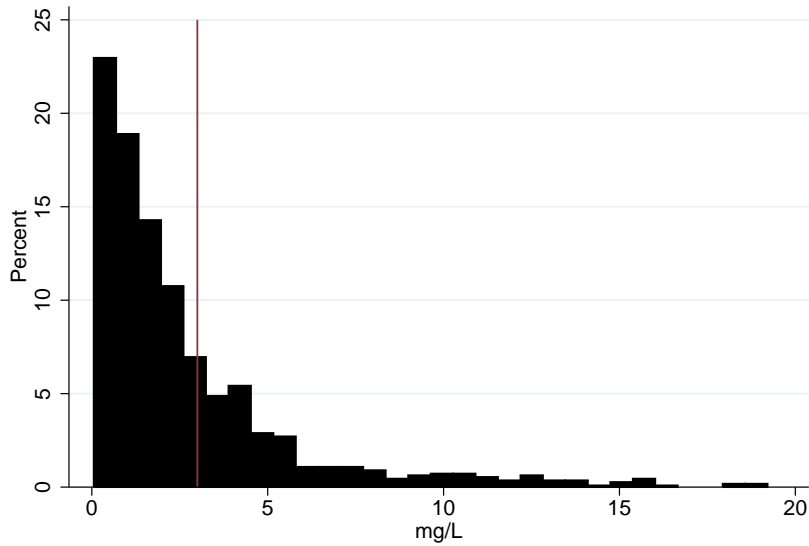
Source: LASI Pilot 2010 Biomarker Sample, N=1,106

Note: Sample is weighted. The cut-off for high cardiovascular risk is shown at 3 mg/L (above is at risk).

Figure 3: Hemoglobin Distribution for Men

Source: LASI Pilot 2010 Biomarker Sample, N= 517

Note: Sample is weighted. The cut-off for anemia among men is shown at 13 g/dL (below is anemic).

11

Figure 4: Hemoglobin Distribution for Women

Source: LASI Pilot 2010 Biomarker Sample, N= 560

Note: Sample is weighted. The cutoff for anemia among women is shown at 12 g/dL (below is anemic).

Figure 5: BMI Distribution for Men

Source: LASI Pilot 2010 Biomarker Sample, N= 545
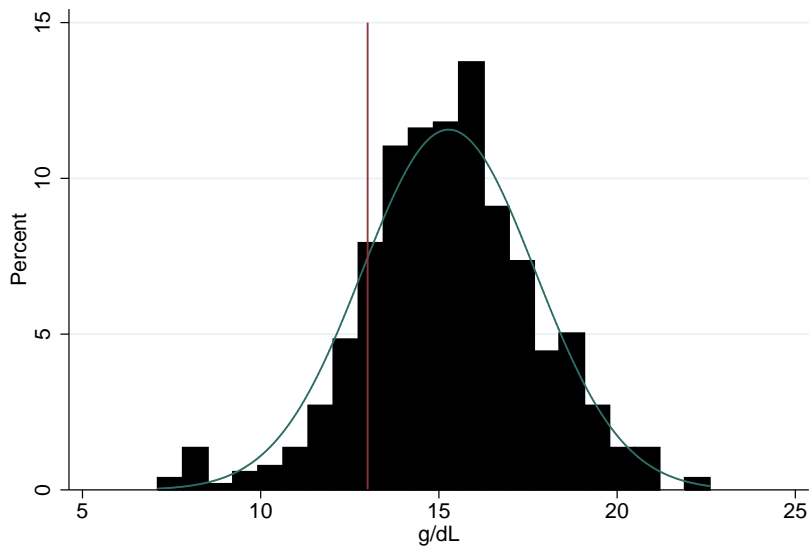
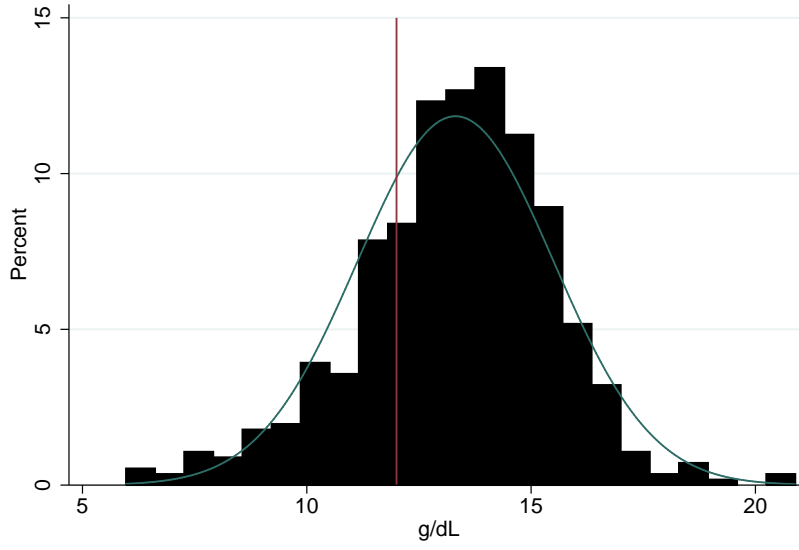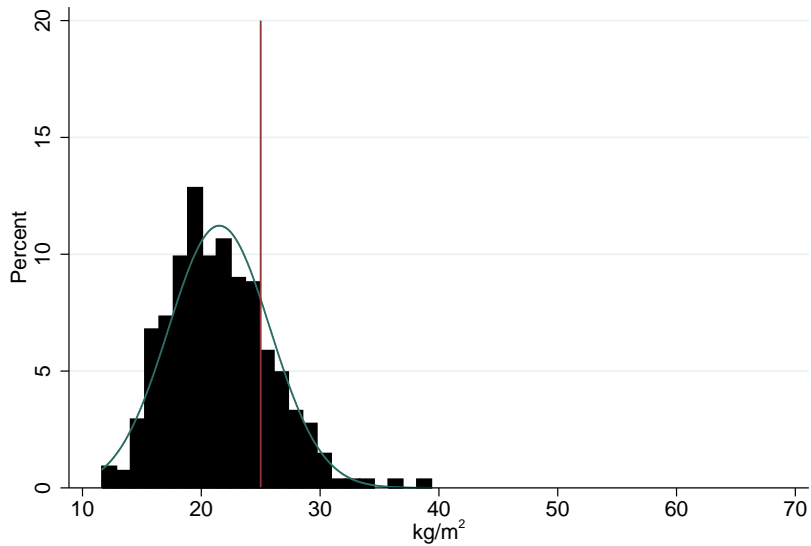Note: Sample is weighted. The cutoff for overweight is shown at 25 kg/m^2.
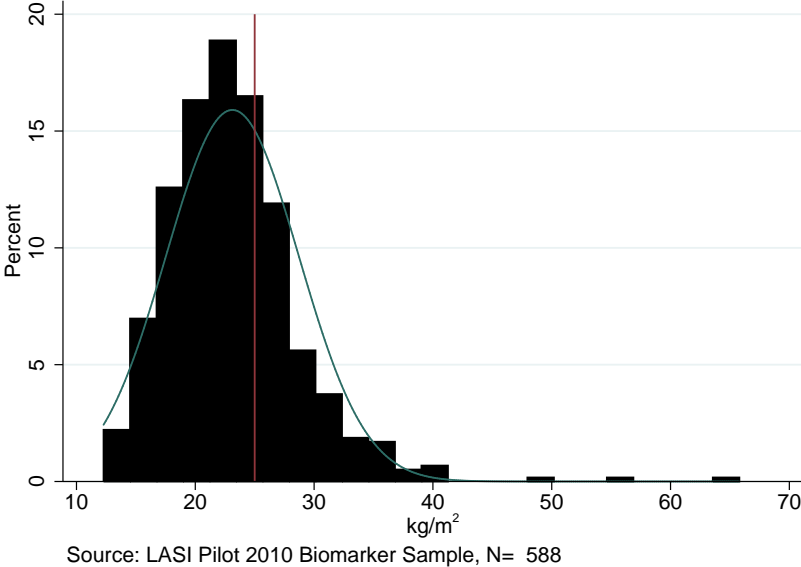
Figure 6: BMI Distribution for Women

Source: LASI Pilot 2010 Biomarker Sample, N= 588

Note: Sample is weighted. The cutoff for overweight is shown at 25 kg/m^2.

13

Table 3: Bivariate Association between Explanatory Variables and Hemoglobin

	Mean	SE	CI	N	t-test P Value
Total	14.23	0.14	[13.96,14.51]	1,077	
Gender					
Male	15.27	0.16	[14.94,15.60]	517	
Female	13.27	0.14	[12.98,13.55]	560	(0.00)
Age Group					
44-54	14.41	0.14	[14.12,14.70]	506	
55-64	14.22	0.25	[13.72,14.73]	295	(0.51)
65-74	13.96	0.24	[13.48,14.44]	185	(0.06)
75+	13.81	0.32	[13.17,14.46]	90	(0.09)
Caste					
Scheduled Caste	13.79	0.25	[13.29,14.30]	179	(0.16)
Scheduled Tribe	13.99	0.32	[13.34,14.65]	137	(0.57)
Other Backward Class	14.54	0.18	[14.18,14.91]	367	(0.16)
None	14.20	0.19	[13.81,14.59]	366	
Education					
Some Schooling	14.88	0.16	[14.56,15.19]	580	
No Schooling	13.55	0.16	[13.22,13.88]	497	(0.00)
Residency					
Urban	14.69	0.33	[14.02,15.36]	257	
Rural	14.09	0.14	[13.80,14.37]	820	(0.10)
State					
Punjab	14.12	0.36	[13.40,14.84]	264	
Rajasthan	13.73	0.24	[13.25,14.21]	275	(0.37)
Kerala	14.69	0.13	[14.43,14.94]	306	(0.14)
Karnataka	14.46	0.29	[13.87,15.05]	232	(0.47)
BMI Group					
Underweight	13.64	0.24	[13.16,14.12]	240	(0.00)
Normal	14.51	0.15	[14.21,14.81]	551	
Overweight	14.49	0.24	[14.01,14.97]	205	(0.92)
Obese	14.00	0.41	[13.18,14.82]	66	(0.18)
Diabetes, Stroke, Hypertension, or Heart Condition					
No	14.11	0.15	[13.81,14.40]	811	
Yes	14.68	0.18	[14.31,15.05]	266	(0.01)
Currently Smokes					
No	14.03	0.13	[13.77,14.30]	907	
Yes	15.21	0.21	[14.79,15.64]	167	(0.00)
Household Living Arrangements					
Married, No Children in Household	14.49	0.28	[13.92,15.07]	147	
Married, Children in Household	14.44	0.15	[14.15,14.74]	709	(0.86)
Not Married, Children in Household	13.48	0.33	[12.82,14.14]	53	(0.01)
Not Married, No Children in Household	13.38	0.18	[13.01,13.75]	168	(0.00)

Note: Sample is weighted. Confidence intervals and T tests account for survey design. Underweight is an indicator for BMI<18.5, normal weight is $18.5 \leq$ BMI<25, overweight is $25 \leq$ BMI<30, and obese is BMI ≥ 30. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease. Hb is measured in g/dL.

Table 4: Bivariate Association between Explanatory Variables and C-Reactive Protein

	Mean	SE	CI	N	t-test P Value
Total	2.69	0.10	[2.49,2.90]	1,106	
Gender					
Male	2.78	0.15	[2.48,3.08]	529	
Female	2.61	0.15	[2.32,2.90]	577	(0.43)
Age Group					
44-54	2.47	0.16	[2.15,2.79]	513	
55-64	2.71	0.16	[2.39,3.03]	310	(0.25)
65-74	2.85	0.22	[2.41,3.30]	194	(0.19)
75+	3.61	0.49	[2.61,4.60]	88	(0.03)
Caste					
Scheduled Caste	2.78	0.23	[2.31,3.25]	180	(0.97)
Scheduled Tribe	2.94	0.41	[2.12,3.76]	130	(0.78)
Other Backward Class	2.48	0.16	[2.16,2.80]	390	(0.30)
None	2.79	0.23	[2.33,3.26]	378	
Education					
Some Schooling	2.53	0.11	[2.30,2.75]	499	
No Schooling	2.88	0.18	[2.52,3.24]	275	(0.10)
Residency					
Urban	3.16	0.19	[2.78,3.53]	290	
Rural	2.53	0.12	[2.30,2.77]	816	(0.01)
State					
Punjab	3.06	0.13	[2.79,3.32]	277	
Rajasthan	3.28	0.26	[2.76,3.80]	268	(0.45)
Kerala	1.96	0.17	[1.63,2.30]	314	(0.00)
Karnataka	2.51	0.15	[2.20,2.81]	247	(0.01)
BMI Group					
Underweight	2.36	0.23	[1.90,2.82]	238	(0.37)
Normal	2.61	0.17	[2.27,2.95]	569	
Overweight	2.92	0.18	[2.56,3.29]	215	(0.25)
Obese	3.77	0.46	[2.85,4.70]	68	(0.02)
Diabetes, Stroke, Hypertension, or Heart Condition					
No	2.61	0.12	[2.37,2.86]	826	
Yes	2.97	0.19	[2.58,3.36]	280	(0.15)
Currently Smokes					
No	2.68	0.11	[2.46,2.90]	938	
Yes	2.71	0.27	[2.17,3.26]	166	(0.90)
Household Living Arrangements					
Married, No Children in Household	2.29	0.21	[1.87,2.71]	152	
Married, Children in Household	2.79	0.11	[2.57,3.02]	729	(0.01)
Not Married, Children in Household	2.08	0.35	[1.37,2.79]	54	(0.61)
Not Married, No Children in Household	2.82	0.28	[2.26,3.39]	171	(0.13)

Note: Sample is weighted. Confidence intervals and t-tests account for survey design. Underweight is an indicator for BMI<18.5, normal weight is $18.5 \leq$ BMI<25, overweight is $25 \leq$ BMI<30, and obese is BMI $\geq$ 30. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease. CRP is measured in mg/L.

Table 3 shows that being female and being underweight are significantly associated with a lower level of Hb, as is a lack of formal schooling. Having a self-reported chronic health condition and smoking are associated with a higher Hb level. Not being married is associated with lower hemoglobin. These results are consistent with a socioeconomic gradient in anemia, and with well-documented gender disparities (Rosero-Bixby and Dow, 2009). For CRP, there are significant bivariate associations (at the 5% level) between the outcome and being age 75 or older (higher for older respondents), rural location (lower for rural respondents), the states of Kerala and Karnataka (lower), being obese (higher), and living with a child while being married (higher). The finding on obesity is consistent with existing evidence on the relationship between inflammation and obesity (Danesh et al., 2004).

These bivariate associations do not adjust for other variables, such as age and state, which are likely to be related to both the outcome and the covariates of interest. We seek to establish whether the findings above remain intact after adjusting for a number of relevant covariates. For example, the relationship between Hb and education may be explained by the fact that earlier, now older, birth cohorts have lower levels of educational attainment. Once we establish the relevance of the key covariates of interest in a multivariate analysis, we decompose CRP differences across groups using the Blinder-Oaxaca approach in order to understand the origin of these differences (Blinder, 1973; Jann, 2008; Liu et al., 2013; Maurer, 2011; Oaxaca, 1973; Powell et al., 2012; Sinning et al., 2008). Although our data do not allow for us to interpret this decomposition in a causal manner, this analysis should still provide us with a preliminary indication as to whether the state level differences in CRP we observe can be explained by differences in endowments of risk factors across states.

We begin by adopting the following regression model:

$$Biomarker_i = \alpha + \beta_1 Gender_i + \beta_2 State_i + \beta_3 Eduation_i + \beta_4 AgeGroup_i \tag{1}$$
$$+\beta_5 Urban_i + \beta_6 Caste_i + \mu_i$$

CRP, Hb, underweight (BMI<18.5), and a combined category for overweight or obese (BMI $\geq$ 25) are modeled as a function of the covariates of interest in a linear regression model (OLS), which is adjusted for weighting and survey design. We use overweight or obese as the outcome instead of obesity as there are relatively few respondents in this category (16 men and 54 women). Our main coefficients of interest

are β_1, β_2, and β_3, reflecting the adjusted association of the outcomes with gender, state, and education. Initially, we do not control for BMI, smoking, living arrangements, or the presence of a self-reported chronic health condition, as these could potentially be outcomes of education or the biomarkers themselves. We then extend the analysis to include these additional variables, and demonstrate that doing so has little effect on our conclusions. We present results for the regression models in the following section. For each outcome, we present both the pooled and gender-stratified analyses.

Given that we observe state-level differences in CRP, we are interested in understanding the source of these differences. We adopt the following decomposition approach, based on linear regression, first proposed by Blinder (1973) and Oaxaca (1973):

$$\overline{CRP}_{State_1} - \overline{CRP}_{State_2} = (\bar{X}_{State_1} - \bar{X}_{State_2})\hat{\beta}_{State_1} \tag{2}$$
$$+\bar{X}_{State_2}(\hat{\beta}_{State_1} - \hat{\beta}_{State_2})$$
$$+(\bar{X}_{State_1} - \bar{X}_{State_2})(\hat{\beta}_{State_1} - \hat{\beta}_{State_2})$$

Conceptually, there are three possible reasons for observing differences in mean CRP levels ($\overline{CRP}$) between two states ($State_1$ and $State_2$): either a difference in the observed covariates (X_{State_1}, X_{State_2}), which is the first term on the right hand side of equation (2), or a difference in the estimated association between the covariates and the outcomes ($\hat{\beta}_{State_1}$, $\hat{\beta}_{State_2}$), which is the second term on the right hand side of equation (2), or an interaction between the two, which is the last term on the right hand side of equation (2) (Daymont and Andrisani, 1984).

We can use this approach to decompose differences in mean CRP levels between State 1 and State 2 into differences in endowments (i.e., establishing what State 2 outcomes would be if $State_2$ had $State_1$'s endowments), differences in the relevant estimated coefficients (i.e., establishing what $State_2$ outcomes would be if $State_2$ had $State_1$'s coefficients), and an interaction between the two. This method relies on the assumption that the state level coefficients are unbiased ($E(\hat{\beta}_{State_1}) = \beta_{State_1}$, $E(\hat{\beta}_{State_2}) = \beta_{State_2}$), and that the state level error terms have mean 0 in expectation ($E(\mu_{State_1}) = 0$, $E(\mu_{State_2}) = 0$). Our data do not allow us to implement a strategy to account for omitted variables, and as a consequence we cannot be certain that this assumption holds in our analysis. Therefore, we are careful to interpret our results with this limitation in mind. Nevertheless, we argue that this analysis is still useful because it provides initial evidence on why the linear predictions for the four states from a standard OLS regression model differ. These findings may be useful for further causal analysis when additional data becomes available.

Table 5: OLS Regression Results for Hemoglobin

Variables	OLS All Hb	OLS All Hb	OLS Men Hb	OLS Men Hb	OLS Women Hb	OLS Women Hb
Age Group: Omitted 45-54						
55-64	-0.314	-0.349	-0.416	-0.531	-0.198	-0.225
	(0.237)	(0.227)	(0.343)	(0.317)	(0.239)	(0.237)
65-74	-0.423*	-0.456**	-0.398	-0.451	-0.409	-0.463
	(0.220)	(0.222)	(0.399)	(0.403)	(0.304)	(0.350)
75+	-0.399	-0.270	-0.852**	-0.869**	0.023	0.175
	(0.287)	(0.260)	(0.364)	(0.383)	(0.407)	(0.335)
Female	-1.969***	-1.880***				
	(0.154)	(0.168)				
Caste: Omitted None of them						
Scheduled caste	0.583	0.612	0.705	0.705	0.474	0.612
	(0.393)	(0.381)	(0.510)	(0.516)	(0.444)	(0.450)
Scheduled tribe	0.436	0.384	0.696*	0.679*	0.163	0.202
	(0.286)	(0.284)	(0.398)	(0.395)	(0.299)	(0.306)
Other backward classs obc	0.030	0.026	0.389	0.412	-0.300	-0.242
	(0.311)	(0.301)	(0.413)	(0.408)	(0.340)	(0.322)
No Formal Education	-0.707***	-0.653**	-0.709*	-0.617	-0.693**	-0.683**
	(0.237)	(0.244)	(0.387)	(0.389)	(0.315)	(0.303)
Rural	-0.455	-0.410	-0.429	-0.458	-0.478	-0.352
	(0.300)	(0.304)	(0.445)	(0.450)	(0.286)	(0.306)
State: Omitted Punjab						
Rajasthan	-0.477	-0.310	-0.356	-0.162	-0.653	-0.534
	(0.417)	(0.419)	(0.463)	(0.479)	(0.509)	(0.491)
Kerala	0.261	0.103	0.385	0.268	0.125	-0.086
	(0.385)	(0.409)	(0.476)	(0.496)	(0.443)	(0.464)
Karnataka	-0.088	-0.044	0.108	0.092	-0.271	-0.206
	(0.419)	(0.432)	(0.495)	(0.506)	(0.462)	(0.481)
BMI Group: Omitted Normal						
Underweight		-0.672***		-0.295		-1.008***
		(0.212)		(0.335)		(0.256)
Overweight		-0.045		0.568		-0.525*
		(0.257)		(0.356)		(0.312)
Obese		-0.018		0.882*		-0.344
		(0.268)		(0.441)		(0.306)
		0.402**		0.358		0.410*
Diabetes, Stroke, Hypert, or Heart		(0.168)		(0.330)		(0.214)
		0.444**		0.509**		0.189
Currently Smokes		(0.219)		(0.252)		(0.350)
Household Living Arrangements Omitted: Married, No Children In HH						
Married, Children in Household		-0.128		-0.448		0.181
		(0.231)		(0.284)		(0.356)
Not Married, Children in Household		-0.354		-1.600		0.269
		(0.321)		(0.999)		(0.429)
Not Married, No Children in Household		-0.253		-0.741*		0.078
		(0.223)		(0.413)		(0.315)
Constant	15.979***	15.371***	15.672***	15.586***	14.322***	13.296***
	(0.470)	(0.554)	(0.576)	(0.676)	(0.539)	(0.619)
Weighted						
Number of Primary Sampling Units (PSU)	63	63	63	63	63	63
Observations	1,048	1,030	505	499	543	531
R-squared	0.223	0.247	0.095	0.138	0.073	0.110

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Sample is weighted. Standard errors are clustered to account for survey design. Underweight is an indicator for BMI<18.5, overweight is 25 ≤ BMI <30, and obese is BMI ≥ 30. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease. Hb is measured in g/dL.

Table 6: OLS Regression Results for Overweight and Underweight

Variables	OLS Men Overweight	OLS Men Overweight	OLS Men Underweight	OLS Men Underweight	OLS Women Overweight	OLS Women Overweight	OLS Women Underweight	OLS Women Underweight
Age Group: Omitted 45-54								
55-64	0.033	0.019	-0.012	-0.016	0.035	0.020	0.008	0.015
	(0.034)	(0.035)	(0.045)	(0.042)	(0.048)	(0.050)	(0.048)	(0.049)
65-74	-0.001	-0.029	0.072	0.093	0.024	0.011	-0.003	0.009
	(0.043)	(0.042)	(0.073)	(0.072)	(0.064)	(0.069)	(0.053)	(0.055)
75+	0.026	-0.004	0.179	0.201**	-0.131***	-0.154***	0.115	0.118
	(0.057)	(0.056)	(0.107)	(0.100)	(0.048)	(0.055)	(0.074)	(0.073)
Caste: Omitted None of them								
Scheduled caste	0.012	-0.002	0.011	0.025	0.017	0.023	0.247**	0.249**
	(0.047)	(0.048)	(0.084)	(0.086)	(0.045)	(0.048)	(0.094)	(0.097)
Scheduled tribe	-0.018	-0.024	-0.125***	-0.128***	0.162***	0.156**	0.003	0.018
	(0.055)	(0.057)	(0.045)	(0.044)	(0.055)	(0.059)	(0.069)	(0.073)
Other backward classs obc	0.085*	0.076	-0.134**	-0.133**	0.119**	0.108*	-0.050	-0.040
	(0.046)	(0.048)	(0.054)	(0.056)	(0.055)	(0.056)	(0.055)	(0.057)
No Formal Education	-0.140***	-0.116**	0.083	0.057	-0.227***	-0.221***	0.113**	0.108**
	(0.042)	(0.045)	(0.079)	(0.081)	(0.057)	(0.060)	(0.042)	(0.048)
Rural	-0.067	-0.063	0.090**	0.080*	-0.076	-0.073	0.102**	0.099**
	(0.049)	(0.044)	(0.042)	(0.040)	(0.052)	(0.051)	(0.050)	(0.045)
State: Omitted Punjab								
Rajasthan	-0.238***	-0.193***	0.241***	0.192***	-0.295***	-0.285***	0.143**	0.128*
	(0.056)	(0.056)	(0.065)	(0.068)	(0.060)	(0.063)	(0.070)	(0.070)
Kerala	-0.211***	-0.198***	0.111	0.054	-0.319***	-0.336***	0.053	0.056
	(0.066)	(0.068)	(0.067)	(0.063)	(0.068)	(0.070)	(0.050)	(0.055)
Karnataka	-0.167**	-0.136**	0.179**	0.122*	-0.304***	-0.311***	0.164***	0.155***
	(0.064)	(0.061)	(0.068)	(0.067)	(0.055)	(0.056)	(0.047)	(0.047)
Diabetes/Stroke/Hypert/Heart		0.135**		-0.080*		0.071		-0.072
		(0.054)		(0.044)		(0.056)		(0.056)
Currently Smokes		-0.089**		0.154***		-0.003		0.041
		(0.040)		(0.055)		(0.114)		(0.152)
Household Living Arrangements Omitted: Married, No Children In HH								
Married, Children in Household		-0.059		0.005		-0.023		0.049
		(0.048)		(0.046)		(0.053)		(0.049)
Not Married, Children in Household		-0.179***		0.253		0.089		0.068
		(0.062)		(0.173)		(0.078)		(0.072)
Not Married, No Children in Household		-0.128*		0.002		-0.016		0.043
		(0.073)		(0.117)		(0.063)		(0.053)
Constant	0.419***	0.455***	0.082	0.098	0.609***	0.615***	-0.044	-0.072
	(0.071)	(0.078)	(0.068)	(0.083)	(0.086)	(0.085)	(0.062)	(0.061)
Weighted	Y	Y	Y	Y	Y	Y	Y	Y
Number of Primary Sampling Units (PSU)	63	63	63	63	63	63	63	63
Observations	532	532	532	532	571	568	571	568
R-squared	0.120	0.158	0.128	0.163	0.170	0.178	0.177	0.184

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Sample is weighted. Standard errors are clustered to account for survey design. Underweight is an indicator for BMI <18.5, overweight is an indicator for BMI ≥ 25. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease.

Table 7: OLS Regression Results for C-Reactive Protein

Variables	OLS All CRP	OLS All CRP	OLS Men CRP	OLS Men CRP	OLS Women CRP	OLS Women CRP
Age Group: Omitted 45-54						
55-64	0.266	0.170	0.170	0.026	0.397	0.336
	(0.215)	(0.205)	(0.362)	(0.335)	(0.283)	(0.269)
65-74	0.421	0.318	0.468	0.500	0.377	0.172
	(0.289)	(0.266)	(0.467)	(0.423)	(0.303)	(0.298)
75+	1.135**	1.194**	0.753	0.740	1.544***	1.712***
	(0.485)	(0.488)	(0.878)	(0.898)	(0.516)	(0.469)
Female	-0.099	-0.139				
	(0.207)	(0.187)				
Caste: Omitted None of them						
Scheduled caste	-0.207	-0.102	-0.039	0.073	-0.534	-0.355
	(0.447)	(0.471)	(0.686)	(0.662)	(0.571)	(0.574)
Scheduled tribe	-0.225	-0.354	-0.646	-0.652	0.148	-0.070
	(0.324)	(0.312)	(0.509)	(0.487)	(0.376)	(0.376)
Other backward classs obc	0.025	0.002	-0.295	-0.167	0.300	0.207
	(0.378)	(0.387)	(0.550)	(0.603)	(0.454)	(0.461)
No Formal Education	-0.216	0.013	0.020	0.228	-0.558	-0.320
	(0.255)	(0.265)	(0.428)	(0.464)	(0.373)	(0.352)
Rural	-0.772***	-0.664***	-0.617**	-0.523*	-0.884**	-0.754**
	(0.209)	(0.227)	(0.292)	(0.290)	(0.379)	(0.352)
State: Omitted Punjab						
Rajasthan	0.472	0.766**	0.331	0.538	0.632	1.016
	(0.329)	(0.380)	(0.450)	(0.470)	(0.577)	(0.628)
Kerala	-1.127***	-0.902***	-1.099***	-0.984**	-1.258***	-0.878**
	(0.246)	(0.274)	(0.318)	(0.400)	(0.333)	(0.357)
Karnataka	-0.451*	-0.168	-0.531	-0.416	-0.346	0.182
	(0.268)	(0.298)	(0.362)	(0.416)	(0.380)	(0.433)
BMI Group: Omitted Normal						
Underweight		-0.506*		-0.189		-0.914***
		(0.279)		(0.360)		(0.318)
Overweight		0.437		0.484		0.332
		(0.282)		(0.533)		(0.315)
Obese		1.212**		2.516*		0.757
		(0.488)		(1.373)		(0.470)
Diabetes, Stroke, Hypert, or Heart		0.534*		0.534		0.485
		(0.289)		(0.426)		(0.374)
Currently Smokes		0.350		0.512		-0.021
		(0.352)		(0.414)		(0.727)
Household Living Arrangements Omitted: Married, No Children in HH						
Married, Children in Household		0.424**		0.300		0.705***
		(0.205)		(0.408)		(0.248)
Not Married, Children in Household		-0.401		-0.382		-0.373
		(0.408)		(0.977)		(0.368)
Not Married, No Children in Household		0.407		0.245		0.610
		(0.340)		(0.755)		(0.380)
Constant	3.557***	2.179***	3.686***	2.585***	3.426***	1.459*
	(0.398)	(0.598)	(0.607)	(0.820)	(0.518)	(0.748)
Weighted	Y	Y	Y	Y	Y	Y
Number of Primary Sampling Units (PSU)	63	63	63	63	63	63
Observations	1,077	1,059	517	510	560	549
R-squared	0.052	0.078	0.048	0.073	0.084	0.125

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Sample is weighted. Standard errors are clustered to account for survey design. Underweight is an indicator for BMI<18.5, overweight is 25≤ BMI <30, and obese is BMI ≥ 30. Diabetes, Stroke, Hypertension or Heart Condition refers to self-reported prior diagnosis of at least one of: diabetes mellitus, stroke, hypertension, and heart disease. CRP is measured in mg/L.

3   Results

Table 5 presents results for the multivariate analysis of Hb. Column 1 shows the results for the pooled sample, controlling for age, gender, state, caste, location, and education. Column 2 adds the following additional variables (smoking, BMI category, living arrangements, and self-reported diagnosis of diabetes, stroke, hypertension, or a heart condition). Columns 3 and 4 replicate these specifications for men, while columns 5 and 6 focus on women only. There is clear evidence for gender and education gradients in the pooled sample, with women having around 2g/dL less Hb than men, and with persons with no formal schooling having about 0.6 g/dL less Hb than those with some education. These results are similar to the bivariate estimates. There are also some gender differences in the coefficients in the stratified model. Older men (ages 75+) have reduced levels of Hb, while being in the obese category is positively associated (at the 10% significance level) for men, while being underweight is negatively associated for women, perhaps reflecting a marker of nutritional status. Smoking is also positively associated with Hb for men, but not for women.

Table 6 shows results for overweight and obese (BMI≥25) and underweight (BMI<18.5). Here, educational gradients are also apparent. For example, men and women with no formal education are between 12 and 23 percentage points less likely to be overweight. Women with no formal education are 11 percentage points more likely to be underweight. Rural residence is also associated with being 8 to 10 percentage points more likely to be underweight. State level differences are also large for both men and women, for example, male residents of Rajasthan are 19 to 24 percentage points less likely to be overweight and 19 to 24 percentage points more likely to be underweight compared to Punjab residents. The equivalent figures for women are 29 to 30 and 13 to 14 percentage points respectively.

Table 7 presents the corresponding analysis for CRP. The overall results are similar to the bivariate analysis. In the pooled sample, the oldest old (those aged 75 and older) have higher CRP levels, while those in rural areas have lower levels, as do those in Kerala. Adding the additional control variables has little effect on the coefficients, although doing so indicates that higher BMI is associated with higher CRP. Controlling for BMI category attenuates the association with urban/rural residence, suggesting that differences in obesity may explain some of the difference in CRP risk between urban and rural areas. Prior diagnosis of diabetes, stroke, hypertension, or a heart condition has a marginally significant positive association with CRP in the pooled sample. Those who are married with a child (or children) living in the same household have

higher CRP than those who are married with no children living in the household. Stratification by gender indicates some differences in the association between the covariates and CRP. For example, the age gradient is only present among women, as is the association with living arrangements. As a robustness check, we implemented binary regression models (logit) for the cut-offs for anemia and cardiovascular risk, and found similar results. Finally, we considered a propensity score matching approach for education, and found similar conclusions regarding an SES gradient in Hb: those with no formal schooling have lower levels of Hb, with associations comparable to those shown in Table 5.

Table 8: Blinder-Oaxaca Decomposition for State Differences in C-Reactive Protein

	Punjab Base Model	Punjab Additional Controls	Kerala Base Model	Kerala Additional Controls	Karnataka Base Model	Karnataka Additional Controls
Rajasthan Mean	3.274***	3.220***	3.274***	3.220***	3.274***	3.220***
	(0.257)	(0.266)	(0.278)	(0.288)	(0.257)	(0.266)
Comparison State Mean	3.055***	3.022***	1.926***	1.932***	2.509***	2.459***
	(0.179)	(0.207)	(0.198)	(0.199)	(0.173)	(0.177)
Difference	0.219	0.198	1.348***	1.288***	0.765**	0.761**
	(0.313)	(0.337)	(0.341)	(0.350)	(0.310)	(0.320)
Endowments	-0.254	0.010	0.346	0.331	-0.353	-0.344
	(0.266)	(0.379)	(0.510)	(0.484)	(0.306)	(0.428)
Coefficients	0.946*	1.246**	1.967**	2.129***	0.799*	0.853*
	(0.467)	(0.531)	(0.717)	(0.736)	(0.430)	(0.473)
Interaction	-0.473	-1.058*	-0.965	-1.171	0.318	0.253
	(0.461)	(0.613)	(0.872)	(0.850)	(0.493)	(0.605)
Additional Controls	N	Y	N	Y	N	Y
	63	63	63	63		
Observations	532	519	560	550	507	500

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Sample is weighted. Standard errors are clustered to account for survey design. The base model column implements the Blinder-Oaxaca decomposition for each state compared to Rajasthan, using the control variables from the specification in column 1 of tables 5 and 7. The additional controls column adds BMI category, smoking, living arrangements, and a diabetes/stroke/hypertension/heart condition. CRP is measured in mg/L.

Finally, we implement the Blinder-Oaxaca decomposition approach to analyze state-level differences in CRP, given that we find substantial disparities in table 7. Table 8 shows the relevant means for the base state (Rajasthan), compared to each of the other states. Column 1 uses the regression coefficients from the base model in tables 5 and 7 above, while column 2 controls for BMI category, smoking, living arrangements and prior diagnosis of diabetes, stroke, hypertension, or a heart condition. The decomposition shows the mean difference, the estimated contribution from endowments, the coefficients, and their interaction (as illustrated in equation 2).

Overall, the difference in mean CRP level is statistically significant between Rajasthan and Kerala, and between Rajasthan and Karnataka. The decomposition indicates that these differences are mainly due to differences in the association between the explanatory variables and CRP, rather than the distribution of risk factor endowments across states. The additional controls do not change this result.

4   Conclusions

Using new data from a representative sample of older Indians, we investigate the relationship between respondents' characteristics and objective measures of their health. We demonstrate the feasibility and value of collecting population-based biomarkers among those aged 45 years and older in a developing country. We find evidence for an education gradient in Hb, but there is no evidence of state-level differences. Despite recent economic growth, the risk of anemia, most likely associated with malnutrition, is higher for women and for those without schooling. Like Hb, there are clear education gradients in underweight and overweight, particularly for women. This is consistent with previous evidence on younger Indians (Subramanian et al., 2009), as well as results on large state differences (Ackerson et al., 2008). We find that about one third of Indians have a CRP level considered to be high risk (>3 mg/L), which is comparable to results from the English Longitudinal Study on Ageing (Hamer and Molloy, 2009).

We also find that CRP is greater among the oldest old and among urban residents. Although there are substantial state-level differences, there is no evidence of an education gradient for CRP, which is consistent with existing evidence from Costa Rica (Rosero-Bixby and Dow, 2009). These findings suggest that CVD is a health problem which is equally likely to affect individuals of all socioeconomic groups in India. When we decompose state-level differences, we find that these disparities are mainly due to differences in the association of risk factors with CRP rather than in the distribution of risk factors.

Kerala has the lowest CRP levels, raising the question of why individuals in Kerala and Karnataka, as opposed to Rajasthan and Punjab, carry less risk of CRP as they grow older. We are cautious about drawing strong conclusions about the relationship between state level characteristics and CRP given that we only have four observations, however the differences in mean CRP levels across states are partly consistent with the hypothesis that CRP risk is negatively associated with economic development, as CRP risk is lowest in Kerala. Kerala has the highest income per capita of the four states in the LASI pilot, 68,930 Rupees per capita per annum, compared to 53,888 Rupees in Punjab, 31,354 in Rajasthan, and 25,272 in Karnataka.

However, on this basis, Karnataka has lower CRP risk than would be expected from its income per capita, as it has the second lowest CRP level. According to the most recent census in 2011, the ranking of the states in terms of other indicators is similar to that as measured by income per capita. For example, life expectancy at birth is 74 in Kerala, 69 in Punjab, 67 in Rajasthan, and 67 in Karnataka. Infant mortality is 12% in Kerala, 30% in Punjab, 52% in Rajasthan, and 35% in Karnataka, while urbanization is 48% in Kerala, 38% in Punjab, 25% in Rajasthan, and 39% in Karnataka. The literacy rate is 94% in Kerala, 76% in Punjab, 66% in Rajasthan and 75% in Karnataka. In terms of religious and ethnic composition, Kerala also has a relatively more equal share of its population from the three main religious categories in India: Hindu, Muslim and Christian. Kerala has relatively few residents in scheduled tribes or scheduled castes, whereas Karnataka, Punjab and Rajasthan have a greater proportion of scheduled caste and scheduled tribes. In summary, Kerala is clearly the most advantaged state of the four included in the LASI pilot, regardless of which measure is used. As we outline above, the fact that it also has the lowest CRP risk is suggestive of a relationship with economic development. Further rounds of the LASI survey which include longitudinal data on all states over time will permit a more comprehensive investigation of the state level correlates of CRP.

The finding that the state-level differences in CRP risk can be explained by differential association of CRP with risk factors suggests that there might be some moderating factors that vary across states. Given that most policies are developed and implemented at state level, this finding is not surprising. In particular, Kerala is the state known for its policies promoting public health and education. Such differences in state policies can contribute to the variation in access to health care and education, and as a follow-up to this decomposition analysis, we therefore further examine heterogeneity in access to health care across states. In fact, we find some support for this hypothesis of state-level policy differences in our data. As shown in figure 7, Kerala has a relatively high percentage of respondents (71%) who report having access to formal health care (a doctor with a recognized qualification), however, Karnataka also has a high percentage of respondents in this category (73%). Therefore, health care access is one factor which could potentially moderate the association between risk factors and inflammation. If this relationship with CRP was found to hold in a larger sample of states, it would indicate the type of policy which could be pursued in order to better target cardiovascular disease in India. This is an important issue which we are unable to fully explore with the current data covering only four states, however this is another key direction for future research.

Figure 7: C-Reactive Protein and Health Care Access by State
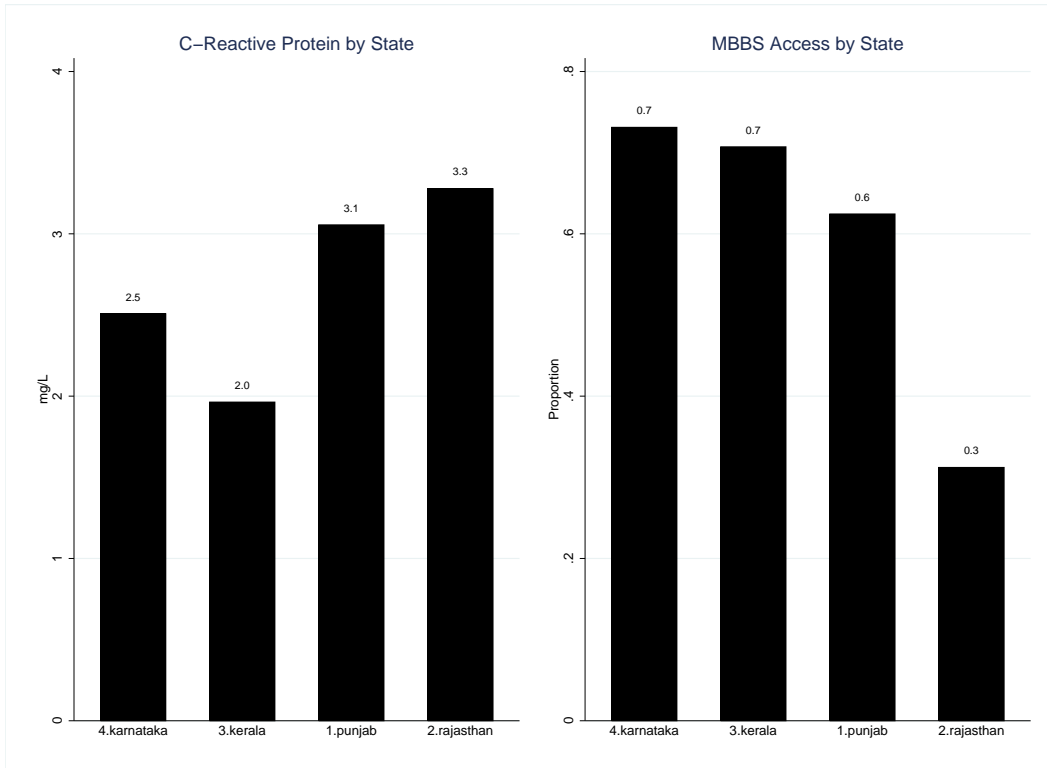
Note: Sample is weighted. The left figure shows the mean CRP level by state, while the right figure shows the proportion of respondents in each state who have ever visited a doctor with a formal qualification (e.g., an M.M.B.S. degree).

There are important limitations to this study. First, we are only able to document associations. Further data are required to establish, for example, whether the observed relationship between Hb and education is causal. In particular, we are careful to interpret the results of the Blinder-Oaxaca decomposition as being indicative only, because our estimates of the association between risk factors and CRP could be affected by omitted variables. Another area where causal analysis would be useful is in further investigating the gender differences we observe, not only in mean level differences, but also in terms of assessing gender differences in the relationship between risk factors and the outcomes of interest. Second, a potentially important question which could be addressed using biomarker data is whether health assessments in survey data which are based on objective data differ from health assessments which are based on subjective data. Overall, biomarkers are likely to provide a useful complement to other health measures when determining the health status of older individuals in lower- and middle-income countries, as well as the health disparities between groups within these countries. The use of biomarkers may help to overcome the drawbacks associated with self-reported health measures.

This measurement question has potential implications for our understanding of health disparities, as well as our understanding of the risk factors for health conditions, such as CVD. Previous studies have found significant differences between objective and subjective measures among older individuals for some outcomes, for example in England using the English Longitudinal Study of Ageing (Johnston et al., 2009), and in India using SAGE (Vellakkal et al., 2013). However, the outcomes we consider in this paper are not well suited to analyzing this question. We do not have self-reported measures of nutrition in the data which we could compare to Hb levels, nor self-reported weight or height which we could use to compare to the objectively measured variables we use in the paper. Although low-grade but persistent inflammation is associated with cardiovascular disease, and CRP is predictive of subsequent cardiovascular events and mortality, CRP is not considered as a direct objective measure of heart disease. Finally, we are presently only able to consider the four pilot states, and it would be interesting to consider the relationships shown in this paper for all states. Future, nationally representative waves of LASI will enable us to examine whether these findings extend to other Indian states, and future iterations of the modules and questionnaires will provide valuable data for answering these and other important research questions.

References

L. K. Ackerson, I. Kawachi, E. M. Barbeau, and S. Subramanian. Geography of underweight and overweight among women in India: a multilevel analysis of 3204 neighborhoods in 26 states. *Economics & Human Biology*, 6(2):264–280, 2008.

V. M. Aguayo, S. Scott, and J. Ross. Sierra Leone - investing in nutrition to reduce poverty: a call for action. *Public Health Nutrition*, 6(07):653–657, 2003.

P. Arokiasamy, D. E. Bloom, J. Lee, K. Feeney, and M. Ozolins. Longitudinal aging study in India: vision, design, implementation, and some early results. In J. P. Smith and M. Majmundar, editors, *Aging in Asia: findings from new and emerging data initiatives*. National Academies Press, 2012.

Y. Balarajan, U. Ramakrishnan, E. Özaltin, A. H. Shankar, and S. Subramanian. Anaemia in low-income and middle-income countries. *The Lancet*, 378(9809):2123–2135, 2011a.

Y. Balarajan, S. Selvaraj, and S. Subramanian. Health care and equity in india. *The Lancet*, 377(9764): 505–515, 2011b.

Y. S. Balarajan, W. W. Fawzi, and S. Subramanian. Changing patterns of social inequalities in anaemia among women in india: cross-sectional study using nationally representative data. *BMJ Open*, 3(3): e002233, 2013.

C. Beghé, A. Wilson, and W. B. Ershler. Prevalence and outcomes of anemia in geriatrics: a systematic review of the literature. *The American Journal of Medicine*, 116(7):3–10, 2004.

H. Beltrán-Sánchez and E. M. Crimmins. Biological risk in the mexican population at the turn of the 21st century. *Journal of Cross-Cultural Gerontology*, 28(3):299–316, 2013.

M. Bentley and P. Griffiths. The burden of anemia among women in india. *European Journal of Clinical Nutrition*, 57(1):52–60, 2003.

A. S. Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455, 1973.

D. E. Bloom, E. T. Cafiero-Fonseca, M. E. McGovern, K. Prettner, A. Stanciole, J. Weiss, S. Bakkila, and L. Rosenberg. The macroeconomic impact of non-communicable diseases in China and India: Estimates, projections, and comparisons. *The Journal of the Economics of Ageing*, 4:100–111, 2014a.

D. E. Bloom, P. Hu, P. Arokiasamy, A. Risbud, T. V. Sekher, S. K. Mohanty, V. Kale, J. O'Brien, C. S. Chien, and J. Lee. Longitudinal Aging Study in India: Biomarker documentation. *RAND Working Paper WR-1043*, 2014b.

J. Bongaarts and Z. Zimmer. Living arrangements of older adults in the developing world an analysis of demographic and health survey household surveys. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 57(3):S145–S157, 2002.

R. Carmel. Anemia and aging: an overview of clinical, diagnostic and biological issues. *Blood Reviews*, 15 (1):9–18, 2001.

P. H. Chaves, R. D. Semba, S. X. Leng, R. C. Woodman, L. Ferrucci, J. M. Guralnik, and L. P. Fried. Impact of anemia and cardiovascular disease on frailty status of community-dwelling older women: the women's health and aging studies I and II. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 60(6):729–735, 2005.

D. Coffey. Early life mortality and height in indian states. *Economics & Human Biology*, Forthcoming, 2014.

E. M. Crimmins. Physiological differences across populations reflecting early life and later life nutritional status and later life risk for chronic disease. *Journal of Population Ageing*, Forthcoming, 2015.

M. Daly. The relationship of c-reactive protein to obesity-related depressive symptoms: A longitudinal study. *Obesity*, 21(2):248–250, 2013.

J. Danesh, J. G. Wheeler, G. M. Hirschfield, S. Eda, G. Eiriksdottir, A. Rumley, G. D. Lowe, M. B. Pepys, and V. Gudnason. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *New England Journal of Medicine*, 350(14):1387–1397, 2004.

T. N. Daymont and P. J. Andrisani. Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources*, 19(3):408–428, 1984.

B. de Benoist, E. McLean, I. Egll, and M. Cogswell. *Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.* World Health Organization, 2008.

A. Deaton and J. Dreze. Poverty and inequality in india: a re-examination. *Economic and Political Weekly*, 37(6):3729–3748, 2002.

A. Deaton and V. Kozel. Data and dogma: the great indian poverty debate. *The World Bank Research Observer*, 20(2):177–199, 2005.

S. D. Denny, M. N. Kuchibhatla, and H. J. Cohen. Impact of anemia on mortality, cognition, and function in community-dwelling elderly. *The American Journal of Medicine*, 119(4):327–334, 2006.

J. Dreze and P. Srinivasan. Widowhood and poverty in rural India: Some inferences from household survey data. *Journal of Development Economics*, 54(2):217–234, 1997.

E. V. Edmonds, K. Mammen, and D. L. Miller. Rearranging the family? income support and elderly living arrangements in a low-income country. *Journal of Human Resources*, 40(1):186–207, 2005.

A. Fernihough and M. E. McGovern. Physical stature decline decline and the health status of the elderly population in england. *Economics & Human Biology*, 16:30–44, 2015.

S. Ghosh. Exploring socioeconomic vulnerability of anaemia among women in eastern indian states. *Journal of Biosocial Science*, 41(06):763–787, 2009.

D. R. Gwatkin. Metrics matter: the case of assessing the importance of non-communicable diseases for the poor. *International Journal of Epidemiology*, 42(5):1211–1214, 2013.

M. Hamer and G. J. Molloy. Association of c-reactive protein and muscle strength in the English Longitudinal Study of Ageing. *Age*, 31(3):171–177, 2009.

Y. Hu and N. Goldman. Mortality differentials by marital status: an international comparison. *Demography*, 27(2):233–250, 1990.

M. E. Hughes and L. J. Waite. Health in household context: Living arrangements and health in late middle age. *Journal of Health and Social Behavior*, 43(1):1, 2002.

B. Jann. A Stata implementation of the Blinder-Oaxaca decomposition. *The Stata Journal*, 8(4):453–479, 2008.

D. W. Johnston, C. Propper, and M. A. Shields. Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, 28(3):540–552, 2009.

P. M. Kearney, M. Whelton, K. Reynolds, P. Muntner, P. K. Whelton, and J. He. Global burden of hypertension: analysis of worldwide data. *The Lancet*, 365(9455):217–223, 2005.

P. Kowal, S. Chatterji, N. Naidoo, R. Biritwum, W. Fan, R. L. Ridaura, T. Maximova, P. Arokiasamy, N. Phaswana-Mafuya, S. Williams, J. J. Snodgrass, N. Minicuci, C. D'Este, K. Peltzer, J. T. Boerma, and the SAGE Collaborators. Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). *International Journal of Epidemiology*, 41(6):1639–1649, 2012.

J. Lee and J. P. Smith. Regional disparities in adult height, educational attainment, and late-life cognition: Findings from the longitudinal aging study in india (LASI). *The Journal of the Economics of Ageing*, 4: 26–34, 2014.

J. Lee, P. Arokiasamy, A. Chandra, P. Hu, J. Liu, and K. Feeney. Markers and drivers: Cardiovascular health of middle-aged and older Indians. In J. P. Smith and M. Majmundar, editors, *Aging in Asia: findings from new and emerging data initiatives*. National Academies Press, 2012.

H. Liu, H. Fang, and Z. Zhao. Urban-rural disparities of child health and nutritional status in China from 1989 to 2006. *Economics & Human Biology*, 11(3):294–309, 2013.

P. Lloyd-Sherlock, J. Beard, N. Minicuci, S. Ebrahim, and S. Chatterji. Hypertension among older adults in low-and middle-income countries: prevalence, awareness and control. *International Journal of Epidemiology*, 43(1):116–128, 2014.

A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet*, 367(9524): 1747–1757, 2006.

J. Maurer. Education and male-female differences in later-life cognition: International evidence from Latin America and the Caribbean. *Demography*, 48(3):915–930, 2011.

T. W. McDade, J. Burhop, and J. Dohnal. High-sensitivity enzyme immunoassay for c-reactive protein in dried blood spots. *Clinical Chemistry*, 50(3):652–654, 2004.

C. J. Murray, T. Vos, R. Lozano, M. Naghavi, A. D. Flaxman, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, and S. Abdalla. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380 (9859):2197–2223, 2013.

G. Myers, N. Rifai, R. Tracy, W. Roberts, R. Alexander, L. Biasucci, J. Catravas, T. Cole, G. Cooper, and B. Khan. CDC/AHA workshop on markers of inflammation and cardiovascular disease-application to

clinical and public health practice-report from the laboratory science discussion group. *Circulation*, 110 (25):E545–E549, 2004.

K. V. Narayan, M. K. Ali, and J. P. Koplan. Global noncommunicable diseases: where worlds meet. *New England Journal of Medicine*, 363(13):1196–1198, 2010.

R. Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3): 693–709, 1973.

S. O'Broin and E. W. Gunter. Screening of folate status with use of dried blood spots on filter paper. *The American Journal of Clinical Nutrition*, 70(3):359–367, 1999.

B. M. Popkin, S. Horton, S. Kim, A. Mahal, and J. Shuigao. Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: The economic costs of the nutrition transition. *Nutrition Reviews*, 59(12):379–390, 2001.

L. M. Powell, R. Wada, R. C. Krauss, and Y. Wang. Ethnic disparities in adolescent body mass index in the united states: the role of parental socioeconomic status and economic contextual factors. *Social Science & Medicine*, 75(3):469–476, 2012.

A. M. Prentice. The emerging epidemic of obesity in developing countries. *International Journal of Epidemiology*, 35(1):93–99, 2006.

M. Rahman. Living arrangements and the health of older persons in developing countries: evidence from rural Bangladesh. *Population Bulletin of the United Nations*, (42-43):20, 2013.

S. I. Rajan and S. Kumar. Living arrangements among Indian elderly: New evidence from national family health survey. *Economic and Political Weekly*, 38(1):75–80, 2003.

M. Ravallion and G. Datt. Why has economic growth been more pro-poor in some states of India than others? *Journal of Development Economics*, 68(2):381–400, 2002.

L. Rosero-Bixby and W. H. Dow. Surprising SES gradients in mortality, health, and biomarkers in a Latin American population of adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 64(1):105–117, 2009.

A. Sen. Health: perception versus observation: Self reported morbidity has severe limitations and can be extremely misleading. *BMJ*, 324(7342):860, 2002.

P. Shetty. Grey matter: ageing in developing countries. *The Lancet*, 379(9823):1285–1287, 2012.

M. Sinning, M. Hahn, and T. K. Bauer. The Blinder-Oaxaca decomposition for nonlinear regression models. *The Stata Journal*, 8(4):480–492, 2008.

S. Subramanian, J. M. Perkins, and K. T. Khan. Do burdens of underweight and overweight coexist among lower socioeconomic groups in India? *The American Journal of Clinical Nutrition*, 90(2):369–376, 2009.

S. Subramanian, D. J. Corsi, M. A. Subramanyam, and G. D. Smith. Jumping the gun: the problematic discourse on socioeconomic status and cardiovascular health in India. *International Journal of Epidemiology*, 42(5):1410–1426, 2013.

S. Vellakkal, S. Subramanian, C. Millett, S. Basu, D. Stuckler, and S. Ebrahim. Socioeconomic inequalities in non-communicable diseases prevalence in India: disparities between self-reported diagnoses and standardized measures. *PloS ONE*, 8(7):e68219, 2013.

N. K. Vikram, A. Misra, M. Dwivedi, R. Sharma, R. Pandey, K. Luthra, A. Chatterjee, V. Dhingra, B. Jailkhani, and K. Talwar. Correlations of c-reactive protein levels with anthropometric profile, percentage of body fat and lipids in healthy adolescents and young adults in urban north India. *Atherosclerosis*, 168(2):305–313, 2003.

World Health Organization. Preventing chronic diseases: a vital investment, 2005.

World Health Organization. India: Country profile, noncommunicable diseases, 2014.