



Patterning of individual heterogeneity in body mass index: evidence from 57 low- and middle-income countries

Rockli Kim¹ · Ichiro Kawachi¹ · Brent Andrew Coull² · Sankaran Venkata Subramanian^{1,3}

Received: 2 October 2017 / Accepted: 12 January 2018 / Published online: 22 January 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Modeling variation at population level has become increasingly valued, but no clear application exists for modeling differential variation in health between individuals within a given population. We applied Goldstein's method (in: Everitt, Howell (eds) *Encyclopedia of statistics in behavioral science*, Wiley, Hoboken, 2005) to model individual heterogeneity in body mass index (BMI) as a function of basic sociodemographic characteristics, each independently and jointly. Our analytic sample consisted of 643,315 non-pregnant women aged 15–49 years pooled from the latest Demographic Health Surveys (rounds V, VI, or VII; years 2005–2014) across 57 low- and middle-income countries. Individual variability in BMI ranged from 9.8 (95% CI: 9.8, 9.9) for the youngest to 23.2 (95% CI: 22.9, 23.5) for the oldest age group; 14.2 (95% CI: 14.1, 14.3) for those with no formal education to 19.7 (95% CI: 19.5, 19.9) for those who have completed higher education; and 13.6 (95% CI: 13.5, 13.7) for the poorest quintile to 20.1 (95% CI: 20.0, 20.2) for the wealthiest quintile group. Moreover, variability in BMI by age was also different for different socioeconomic groups. Empirically testing the fundamental assumption of constant variance and identifying groups with systematically large differentials in health experiences have important implications for reducing health disparity.

Keywords Heterogeneity · Body mass index · Health inequalities · Variation · Low and middle income countries

Introduction

Compared to conventional epidemiologic risk factor research estimating probabilistic average associations at a single level, multilevel modeling accommodates complexity in contextual heterogeneity [1–4]. Hence, multilevel modeling has become increasingly common in health literature that aims to promote population health and health equity. To date, multilevel studies have been restricted to

modeling variation at higher level(s) only (i.e., between-population), while assuming a single constant variance at the individual level (i.e., within-population). The differential variation in health between individuals within a population remains unstudied [5, 6].

The fundamental assumption of homogeneity is quite unrealistic for many physiological, behavioral, and social outcomes [7]. People of different age, gender, race/ethnicity, and socioeconomic status may be healthier or sicker on average, and also differently variable in terms of their health status. An alternative approach proposed by Goldstein [8] recognizes and explicitly models such “complex level-1 variation” as a function of specified predictor variable(s) [8]. While the technical advantages of this methodology have been discussed previously [7, 8], its substantive importance has been underappreciated and it has not been applied in health literature. Identifying groups with particularly large inequality in health experiences and understanding factors that systematically contribute to differential variation have important implications for individual and population health [5, 9].

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10654-018-0355-2>) contains supplementary material, which is available to authorized users.

✉ Sankaran Venkata Subramanian
svsubram@hsph.harvard.edu

¹ Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, USA

² Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

³ Population Health and Geography, Harvard Center for Population and Development Studies, 9 Bow Street, Cambridge, MA 02138, USA

Increasing within-group inequality in body mass index (BMI) has been consistently documented in different populations [10–12]. This paper aims to explicitly model individual heterogeneity in women's BMI using global data pooled from 57 low- and middle-income countries (LMICs). In this novel application, we specifically investigate the following three questions. First, we explore whether women of different sociodemographic characteristics are more or less variable in terms of their BMI measures. For instance, there are physiological reasons to anticipate that older women, on average, have higher BMI than younger women, but it is also important to quantify whether variability in BMI also systematically depends on age. Second, we attempt to model the full complexity of inter-individual variation in BMI as a function of a combination of sociodemographic variables. Lastly, to better understand the stochastic and systematic components of individual variation in BMI, we examine the proportion of variance explained by basic socioeconomic characteristics, each individually and collectively, for different age groups.

Methods

Data source and sampling plan

The data for this study were extracted from the latest cross-sectional Demographic and Health Surveys (DHS) conducted in 57 LMICs between 2005 and 2014. DHS includes standardized and representative sampling of participants, objective measurement of anthropometric measures, and high response rates [13]. Given the standardized data collection procedures across countries and consistent content over time, DHS allows comparability across populations cross-sectionally and over time [13]. DHS collected individual observations following a probability-based cluster sampling procedure, which was then adapted to specific contexts within each country. Sampling frames were first developed on the basis of non-overlapping units of geography (identified as the primary sampling units (PSUs)) that cover the entire country and a fixed proportion of households were selected using systematic sampling within each PSU [14].

Study population and sample size

Our analytic study population was pooled from the DHS rounds V, VI, or VII, whichever was the latest round for each of the 57 countries. Of the 912,444 women, 4801 women younger than 15 or older than 49 years, 68,571 women who were pregnant at the time of the survey, and 175,380 who were not measured for height and/or weight by the study protocol did not meet the eligibility criteria for the analysis.

Moreover, 19,891 women were excluded for missing anthropometric measures and 322 for having biologically implausible height (< 100 or > 200 cm) and/or weight (< 20 or > 200 kg) measures. Lastly, 164 women were excluded for missing information on education level and marital status, leaving 643,315 women in the final analytic sample (Fig. 1).

Outcome

BMI (kg/m^2) was calculated as weight (kg) divided by the square of height (m^2). Trained investigators weighed each woman using a solar-powered scale with an accuracy of 100 g and measured height using an adjustable board calibrated in millimeters.

Explanatory variables

The following five sociodemographic variables, including age, place of residence, household wealth, education, and marital status, were considered in the analysis. Women's age was categorized into seven groups of 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, and 45–49 years. A binary variable for place of residence was defined as census-based urban versus rural. In DHS, household wealth was captured through a composite index of relative standard of living derived from country-specific indicators of asset ownership, housing characteristics and water and sanitation facilities, which was then divided into quintiles for each country [15]. Women's education was coded in four categories indicating no schooling, primary schooling, secondary schooling, and higher schooling. Finally, a variable with three categories of never married, married/living together, divorced/separated/widowed was used for women's marital status.

Statistical analysis

The individual files from 57 countries were combined to create a pooled dataset. First, a linear regression model for BMI, adjusting for all the pre-specified covariates (women's age, place of residence, marital status, education, and wealth), was constructed to serve as the base model for comparison with subsequent models (Model 1). In this ordinary least squares model assuming homoscedasticity, the level-1 (between-individual) variance in BMI is estimated as $e_0 \sim \text{i.i.d.} \mathcal{N}(0, \sigma_{e_0}^2)$. Next, we applied Goldstein's method [8] to model BMI variance as a function of age (Model 2a), type of residence (Model 2b), wealth quintiles (Model 2c), education (Model 2d), and marital status (Model 2e). At level-1, the residuals are no longer summarized with one variance, and instead a variance covariance matrix is estimated. All the calculations used to derive the variances for Models 2a to e are described in detail in

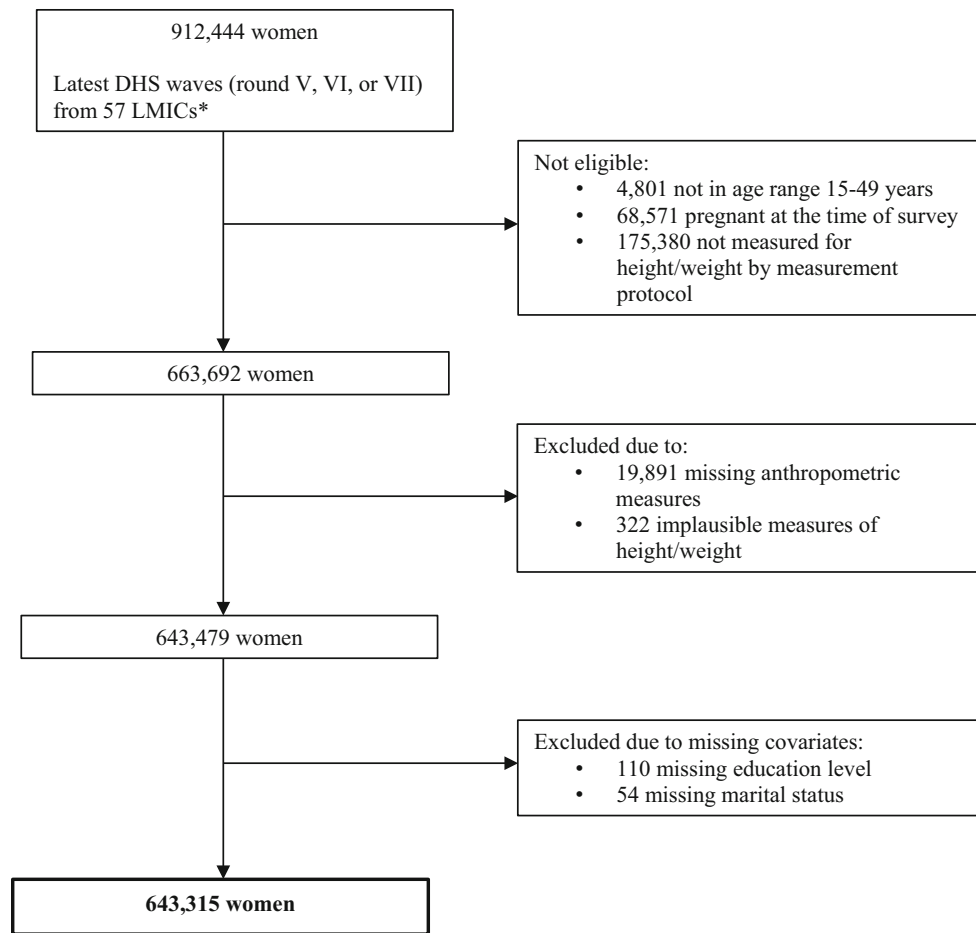


Fig. 1 Final analytic sample from 57 low and middle income countries (Demographic Health Surveys, 2005–2014). *Albania (2008–2009), Armenia (2005), Azerbaijan (2006), Bangladesh (2014), Benin (2011–2012), Bolivia (2008), Burkina Faso (2010), Burundi (2010), Cambodia (2014), Cameroon (2011), Chad (2014–2015), Colombia (2010), Comoros (2012), Congo (Brazzaville) (2011–2012), Congo (DRC) (2013–2014), Cote d’Ivoire (2011–2012), Dominican Republic (2013), Egypt (2014), Ethiopia (2011), Gabon (2012), Gambia (2013), Ghana (2014), Guinea (2012), Guyana (2009), Haiti (2012), Honduras (2011–2012), India

(2005–2006), Jordan (2012), Kenya (2014), Kyrgyz Republic (2012), Lesotho (2014), Liberia (2013), Madagascar (2008–2009), Malawi (2010), Maldives (2009), Mali (2012–2013), Moldova (2005), Mozambique (2011), Namibia (2013), Nepal (2011), Niger (2012), Nigeria (2013), Pakistan (2012–2013), Peru (2012), Rwanda (2014), Sao Tome and Principe (2008–2009), Senegal (2010–2011), Sierra Leone (2013), Swaziland (2006–2007), Tajikistan (2012), Tanzania (2010), Timor-Leste (2009–2010), Togo (2013–2014), Uganda (2011), Yemen (2013), Zambia (2013–2014), Zimbabwe (2010–2011)

Technical Appendix. To check for consistency in the patterning of BMI variation by education and by wealth, we also conducted country-specific analyses for Models 2c and d. To explore potential interaction between age and socioeconomic factors, we first included fixed effects of interaction between age and education (Model 3a). Then, we further modeled BMI variance as a function of all possible cross-classification of age groups and education levels (Model 3b). This procedure was repeated for the interaction between age and wealth quintiles (Model 4a, b; more details presented in Technical Appendix). Lastly, we performed a stratified analysis by age groups to test whether differential amount of BMI variance gets explained within these subgroups after adjusting for each of the socioeconomic predictors individually and all collectively.

The proportion of variance explained by covariate adjustments were computed by subtracting the variance of the model with more terms from the variance of the initial model, and converting to percentage. The AIC from different models were compared for model fit [16]. All analyses were adjusted for country fixed effects and PSU random effects. Multilevel modeling was performed using MLwiN 2.32.

Results

Of the 643,315 women included in the final analytic sample, the average BMI was 23.2 kg/m² with a standard deviation (SD) of 5.1 kg/m². Both mean and dispersion in

BMI varied by sociodemographic predictors, such that older women had higher mean and SD compared to younger women (mean: 25.5 kg/m², SD: 5.9 kg/m² vs. mean: 20.9 kg/m², SD: 3.4 kg/m²). The SD in BMI was also larger for women of higher education and wealthier quintiles (Table 1). In the multivariable linear regression model, the predictor variables were on average associated with BMI in the expected direction (Supplementary Table 1).

Assuming $e_0 \sim \text{i.i.d.}N(0, \sigma_{e_0}^2)$ in Model 1, the variance in BMI at individual level was 16.7 (95% CI: 16.6, 16.7) (Table 2). However, results from Models 2a to e indicate that individual residuals in BMI were *not* independent of each of the sociodemographic variables. The BMI variance within the oldest age group was approximately 2.5 times larger than variance within the youngest group: 9.8 (95% CI: 9.8, 9.9) for women aged 15–19 years to 23.2 (95% CI: 22.9, 23.5) for 45–49 year olds. Women were much more variable in terms of BMI in urban areas (variance estimate (VE): 20.2; 95% CI: 20.1, 20.3) compared to rural areas (VE: 13.9; 95% CI: 13.8, 13.9). By education level, BMI

was found to be the least variable among women with no formal education (VE: 14.2; 95% CI: 14.1, 14.3) and the most variable among those who have completed higher education (VE: 19.7; 95% CI: 19.5, 19.9). A similar pattern was observed for wealth quintiles, with the VE being substantially smaller for the poorest quintile group (VE: 13.6; 95% CI: 13.5, 13.7) compared to the wealthiest group (VE: 20.1; 95% CI: 20.0, 20.2). In terms of marital status, the widowed/divorced/separated women had VE of 20.8 (95% CI: 20.6, 21.1), which was much larger than the VE for never married women (VE: 12.5; 95% CI: 12.4, 12.6) (Table 2). The pattern of increasing variation in women's BMI according to education and wealth was consistently found in country-specific analyses, with some differences in the magnitude (Supplementary Table 2).

On average, there was a significant interaction between age and socioeconomic factors, with the positive association between age and BMI being stronger for more educated and wealthier women (Supplementary Table 3). Independent of such interactive effects in the fixed part, residual variance in BMI was also found to be different for

Table 1 Mean BMI and standard deviation (SD) by individual predictors

Predictors	N	BMI (kg/m ²)				
		Mean	SD	5th	50th	95th
Age						
15–19	127,048	20.9	3.4	16.3	20.3	26.7
20–24	107,217	22.0	4.0	16.8	21.3	29.4
25–29	104,225	23.1	4.7	17.0	22.2	31.9
30–34	90,212	24.1	5.1	17.2	23.1	33.6
35–39	82,537	24.6	5.4	17.2	23.7	34.6
40–44	70,920	25.2	5.8	17.3	24.4	35.8
45–49	61,156	25.5	5.9	17.4	24.7	36.1
Type of residence						
Urban	283,827	24.3	5.4	17.5	23.6	34.5
Rural	359,488	22.2	4.6	16.6	21.3	31.2
Level of education						
No education	154,950	21.9	4.5	16.3	20.8	30.3
Primary	172,811	23.4	5.0	17.1	22.4	32.9
Secondary	251,918	23.6	5.2	17.1	22.7	33.6
Higher	63,636	24.9	5.1	18.0	24.1	34.3
Wealth quintiles						
Lowest	116,378	22.3	4.6	16.4	21.0	30.7
Secondary	120,233	22.8	4.9	16.6	21.5	31.9
Middle	125,815	23.1	5.0	16.8	22.0	32.6
Fourth	133,638	23.6	5.2	17.1	22.7	33.5
Highest	147,251	24.3	5.3	17.6	23.6	34.2
Marital status						
Never married	169,129	21.5	4.0	16.5	20.9	28.8
Married/living together	419,031	23.9	5.2	17.0	22.7	33.6
Widowed/divorced/separated	55,155	24.3	5.4	17.3	23.3	34.2

Table 2 Individual level variance estimates (VE) in BMI modeled as a constant and as a function of each of the selected sociodemographic characteristic

Random	Model 1: Constant		Model 2a: Age heterogeneity		Model 2b: Type of residency heterogeneity		Model 2c: Education heterogeneity		Model 2d: Wealth heterogeneity		Model 2e: Marital status heterogeneity	
	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI
Level-1 VE	16.67	(16.61, 16.73)										
Age												
15–19 years			9.84	(9.76, 9.92)								
20–24 years			12.95	(12.84, 13.06)								
25–29 years			16.34	(16.20, 16.49)								
30–34 years			18.82	(18.64, 18.99)								
35–39 years			20.50	(20.30, 20.70)								
40–44 years			22.79	(22.55, 23.03)								
45–49 years			23.21	(22.94, 23.47)								
Type of residency												
Urban					20.21	(20.10, 20.32)						
Rural					13.88	(13.81, 13.94)						
Level of education												
No education							14.20	(14.09, 14.30)				
Primary education							16.57	(16.45, 16.68)				
Secondary education							17.51	(17.41, 17.60)				
Higher education							19.69	(19.46, 19.91)				
Wealth quintiles												
Q1									13.62	(13.51, 13.73)		
Q2									14.85	(14.73, 14.97)		
Q3									15.92	(15.79, 16.05)		
Q4									17.93	(17.78, 18.06)		
Q5									20.10	(19.95, 20.25)		
Marital status												
Never married											12.50	(12.42, 12.59)
Married/living together											17.81	(17.73, 17.88)

Table 2 (continued)

Random	Model 1: Constant		Model 2a: Age heterogeneity		Model 2b: Type of residency heterogeneity		Model 2c: Education heterogeneity		Model 2d: Wealth heterogeneity		Model 2e: Marital status heterogeneity	
	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI	VE	95% CI
Widowed/divorced/ separated											20.84	(20.59, 21.09)

Adjusted for all covariates (women's age, place of residence, education, wealth, marital status), country fixed effects, and PSU random effects

different combination of sociodemographic groups. Even within the same age group, BMI variance was much larger for more educated and wealthier groups. For instance, while the overall variance for 15–19 year olds was 9.8 (95% CI: 9.8, 9.9), the VE ranged from 7.7 (95% CI: 7.5, 7.9) for 15–19 year olds with no formal education to 11.9 (95% CI: 11.4, 12.4) for those who have completed higher education. Among 45–49 year old women, the variance was smallest for the least educated group (VE: 19.1; 95% CI: 18.7, 19.4) and largest for the secondary education group (VE: 27.6; 95% CI: 26.9, 28.1) (Fig. 2a). A similar trend was observed by wealth quintiles. Within the youngest age group, the VE ranged from 8.3 (95% CI: 8.2, 8.5) for the poorest quintile to 11.7 (95% CI: 11.5, 11.9) for the wealthiest quintile. Within the oldest age group, the VE ranged from 19.3 (95% CI: 18.8, 19.8) for the poorest quintile to 26.6 (95% CI: 26.0, 27.3) for the wealthiest quintile (Fig. 2b). The exact VEs from Models 3b and 4b are presented in Supplementary Table 4.

Lastly, adjusting for type of residence, education, wealth, and marital status explained less than 1% of the individual variation in BMI for all age groups. The explanatory variables that had statistically significant association with BMI on average had almost no systematic contribution to individual variation, and this remained consistent across all age groups (Table 3). As expected, the fixed effect estimates were largely unaffected by modeling level-1 heterogeneity, but the model fit as measured by AIC was better when BMI variance was modeled as a function of the selected predictors (Supplementary Table 5).

Discussion

In this global analysis pooling the latest cross-sectional data from 57 LMICs, we provide an empirical evidence that assuming individual variation in BMI to be constant erroneously masks the underlying systematic differences in variability by sociodemographic characteristics. Women of older age, living in urban areas, more educated, wealthier,

and widowed/divorced/separated groups exhibited higher variability in respect to their BMI measures compared to their reference groups. Existing literature on the association between socioeconomic status and BMI [17] should be interpreted with the understanding that average association corresponds to abstractions that do not correctly represent individual heterogeneity. Modeling variance in health outcomes directly as a function of explanatory variables and quantifying the extent to which individuals differ more from each other in some groups than in other groups can uncover important insights into how different types of people vary among themselves [8, 18].

The technical advantages of modeling complex level-1 variation have been discussed previously. Very briefly, heterogeneous models are recommended to obtain unbiased estimates of standard errors for the fixed-effect and random-effect parameters for any data that violate the homogeneity assumption [19] because the validity of inferences about model parameters may be severely affected [20, 21]. In multilevel analysis, inappropriate assumption of homoscedasticity at level-1 can lead to exaggerated estimates of contextual differences as heterogeneity between-individuals can confound between-context variation [7, 8]. In our analyses, we found that modeling the complex level-1 variation indeed improved the statistical fit of all models.

The observed pattern in BMI heterogeneity by socioeconomic predictors in our analysis was somewhat contrary to what was found in the context of the United States, where the increase in BMI dispersion was shown to be greater for the disadvantageous groups of non-Hispanic blacks and individuals with less than a high school education and high school graduates [10]. According to the “Schmalhausen’s Law”, under severe or unusual stress conditions, even small environmental and genetic differences may lead to major effect in manifested physiology and development [22]. Assuming that predictors such as education and wealth are a reasonable proxy for an individual’s social condition, BMI was indeed shown to be more sensitive to random disturbances caused by more stressful conditions in the US, but more variable in the

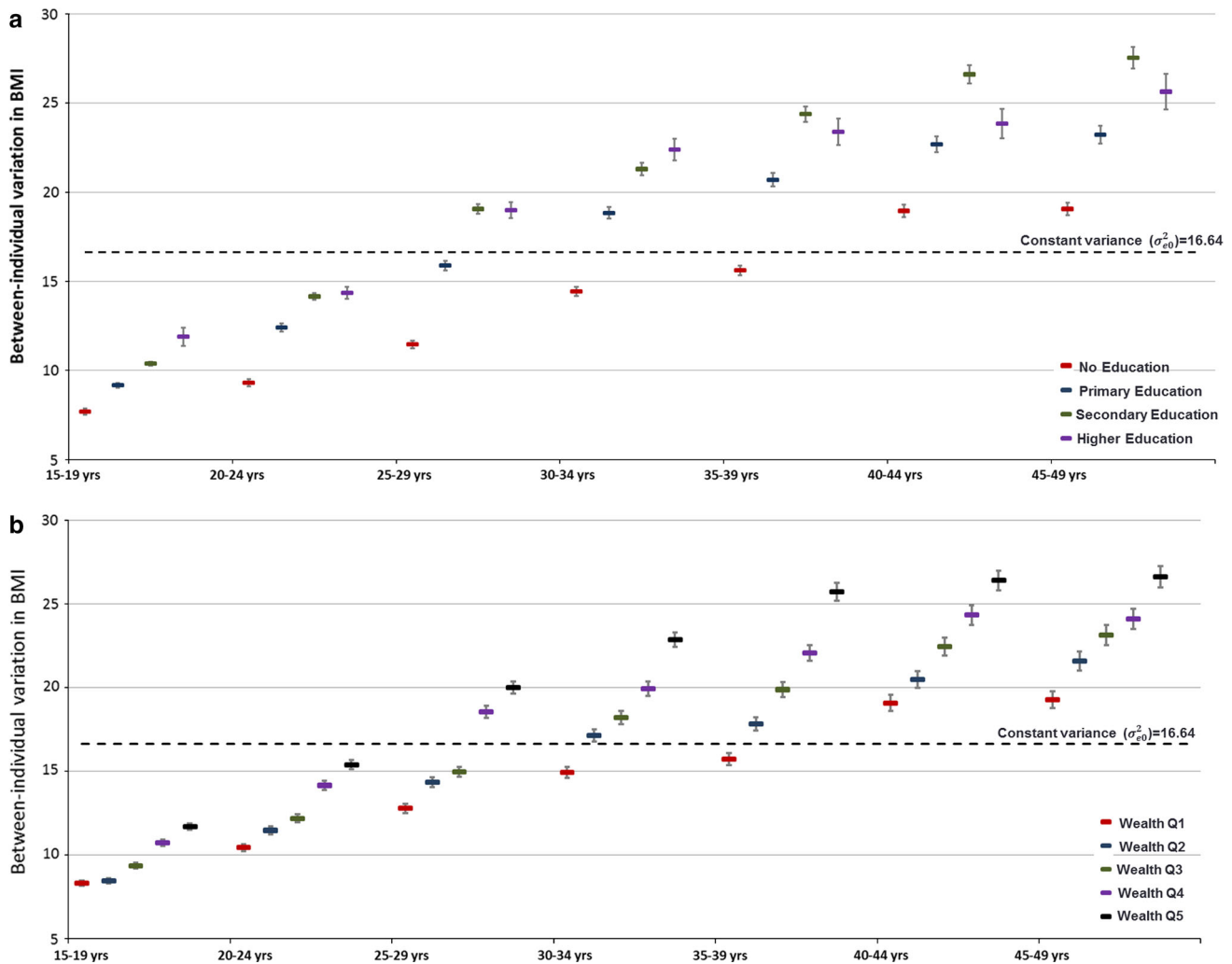


Fig. 2 Differential variability in individual BMI (95% CI) by **a** women's age and education, **b** women's age and wealth quintiles. Adjusted for all covariates (women's age, place of residence,

education, wealth, marital status), country fixed effects, and PSU random effects, and fixed effect for interaction between **a** age and education, **b** age and wealth quintiles

better-off conditions in the LMICs. This reinforces the need to further examine how diverse aspects of the social and physical environments interact with individual characteristics or susceptibilities and result in varied effects in different contexts [23].

We explored two possible explanations for the presence of complex variation in BMI. First, the same set of factors may have differential effects on individuals, also known as *treatment heterogeneity* in clinical settings [24]. Interaction effect is routinely tested in the fixed part of the model when the same exposure-outcome relationship is, on average, expected to be different for particular subgroups. Our findings suggest that independent of such average interactive effects there may be important interaction in variability as well. For instance, in addition to BMI variance being larger for older women, the observed age heterogeneity in BMI was also systematically patterned by

education level and wealth quintiles. Inequality in BMI within the oldest and the wealthiest (or most educated) group was more than three times larger than the youngest and the poorest (or least educated) group. The increase in BMI heterogeneity by age may be explained by differential long-term weight gain that results from a combination of accumulated effects of obesogenic behaviors, including changes in the consumption of specific foods and beverages and amount of physical activity, as well as biological changes over lifecourse [25, 26]. The consistent social gradient in the BMI heterogeneity by age further supports that some health behaviors or proclivities that affect long-term weight are likely conditioned by some common social processes, including childhood conditions [27].

The second explanation for the observed complexity in BMI heterogeneity may be attributed to a completely different set of factors driving the variation within different

Table 3 Variance estimate (VE) stratified by seven age groups and proportion explained (%) by basic socioeconomic factors independently and jointly

	15–19 years N = 127,048	20–24 years N = 107,217	25–29 years N = 104,225	30–34 years N = 90,212	35–39 years N = 82,537	40–44 years N = 70,920	45–49 years N = 61,156
Null							
VE (95% CI)	9.69 (9.61, 9.77)	12.85 (12.74, 12.96)	16.26 (16.11, 16.40)	18.55 (18.37, 18.73)	20.14 (19.94, 20.35)	22.06 (21.82, 22.30)	22.06 (21.80, 22.32)
Type of residency							
VE (95% CI)	9.69 (9.61, 9.77)	12.86 (12.75, 12.97)	16.28 (16.14, 16.43)	18.59 (18.41, 18.76)	20.17 (19.97, 20.37)	22.07 (21.83, 22.31)	22.10 (21.84, 22.36)
% explained	0.0	- 0.1	- 0.2	- 0.2	- 0.1	0.0	- 0.2
Education							
VE (95% CI)	9.69 (9.61, 9.77)	12.86 (12.74, 12.97)	16.29 (16.14, 16.43)	18.59 (18.41, 18.77)	20.13 (19.93, 20.33)	22.07 (21.83, 22.31)	22.10 (21.84, 22.36)
% explained	0.0	0.0	- 0.2	- 0.2	0.1	0.0	- 0.2
Wealth							
VE (95% CI)	9.68 (9.60, 9.76)	12.84 (12.73, 12.96)	16.24 (16.09, 16.38)	18.53 (18.35, 18.70)	19.97 (19.77, 20.17)	21.92 (21.68, 22.15)	21.92 (21.66, 22.17)
% explained	0.1	0.0	0.1	0.1	0.9	0.7	0.7
Marital status							
VE (95% CI)	9.68 (9.60, 9.76)	12.85 (12.74, 12.96)	16.26 (16.11, 16.40)	18.55 (18.37, 18.72)	20.15 (19.94, 20.35)	22.06 (21.82, 22.30)	22.05 (21.80, 22.31)
% explained	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Full adjustment							
VE (95% CI)	9.66 (9.58, 9.74)	12.85 (12.73, 12.96)	16.24 (16.10, 16.38)	18.51 (18.33, 18.69)	19.96 (19.76, 20.16)	21.90 (21.66, 22.13)	21.90 (21.64, 22.15)
% explained	0.3	0.0	0.1	0.2	0.9	0.7	0.7
Adjusted for country fixed effects and PSU random effects							

segments of the population. In this case, in-depth substantive expertise knowledge is necessary to conduct stratified analyses by meaningfully defined subgroups with potentially different sets of risk factors relevant for each. We attempted to explore the stochastic and systematic components of the within-group variation by different age groups, but found little support for differential amount of variation being explained across the age groups. A recent study from Indonesia found that BMI in men tend to become more variable over time compared to women, and that socioeconomic factors explained more of the variation in BMI among men [12]. Due to DHS data on anthropometric measurements being restricted to reproductive age women only in most countries, we were not able to test for gender-specific determinants of BMI heterogeneity at a global scale. For India, a country where BMI measure was collected for 64,958 men aged 15–49 years, we found that socioeconomic factors explain 3.5% of the variation in men and 1.1% in women (data not shown). Future investigations using a larger set of predictors can potentially identify determinants of inter-individual variation that are specific to each of the subgroups that exhibit substantially larger inequality.

While we have pooled our global analytic sample from nationally representative surveys across a broad range of countries, the cross-sectional nature of our dataset inhibits exploration of potential changes in individual variation over time. A study based on repeated cross-sectional data in Indonesia has shown increasing within-group inequalities in BMI over time (1993–2008) that was greatest among individuals in low education and low per capita expenditure groups [12]. Whether this trend holds true across other LMICs should be tested. Additionally, our study population was restricted to young- and middle-aged women. However, we found the same pattern in BMI heterogeneity by sociodemographic factors among a sample of men in India where BMI measure was available (Supplementary Table 6). Lastly, our analysis was limited to a small number of predictor variables that were consistently measured across the countries, but future studies should explore data with diverse sociodemographic, dietary consumption, and genetic information to provide more comprehensive evidence on individual heterogeneity in BMI as well as their relative contributions to variously defined subgroups.

In conclusion, our study suggests that individual heterogeneity should be routinely questioned and empirically tested rather than being treated as “error” or “chance” phenomenon that can be averaged away [28]. The nonrandom patterns of variability observed in BMI suggest that examination of the degree and pattern of heterogeneity between individuals within a population may provide information not evident from the analysis of mean values alone [29]. This approach is especially pertinent

given the increasing within-population inequalities for many health outcomes and related risk factors [30, 31]. Identifying groups with systematically large variation and understanding their specific determinants are important for preventive strategies that aim to reduce health disparity [32].

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical review The study was reviewed by Harvard T. H. Chan School of Public Health Institutional Review Board and was considered exempt from full review because the study was based on an anonymous public use data set with no identifiable information on the study participants.

References

1. Subramanian S. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med.* 2004;58(10):1961–7.
2. Frieden TR. A framework for public health action: the health impact pyramid. *Am J Public Health.* 2010;100(4):590–5.
3. Marmot M, Friel S, Bell R, Houweling TA, Taylor S. Health CoSDo. Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet.* 2008;372(9650):1661–9.
4. Arcaya MC, Tucker-Seeley RD, Kim R, Schnake-Mahl A, So M, Subramanian S. Research on neighborhood effects on health in the United States: a systematic review of study characteristics. *Soc Sci Med.* 2016;168:16–29.
5. Merlo J. Invited commentary: multilevel analysis of individual heterogeneity—a fundamental critique of the current probabilistic risk factor epidemiology. *Am J Epidemiol.* 2014;180(2):208–12.
6. Keyes K, Galea S. What matters most: quantifying an epidemiology of consequence. *Ann Epidemiol.* 2015;25(5):305–11.
7. Browne WJ, Draper D, Goldstein H, Rasbash J. Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Comput Stat Data Anal.* 2002;39(2):203–25.
8. Goldstein H. Heteroscedasticity and complex variation. In: Everitt B, Howell D, editors. *Encyclopedia of statistics in behavioral science.* Hoboken: Wiley; 2005. p. 223–32.
9. Razak F, Smith GD, Subramanian S. The idea of uniform change: is it time to revisit a central tenet of Rose’s “Strategy of Preventive Medicine”? *Am J Clin Nutr.* 2016;104(6):1497–507.
10. Krishna A, Razak F, Lebel A, Smith GD, Subramanian S. Trends in group inequalities and interindividual inequalities in BMI in the United States, 1993–2012. *Am J Clin Nutr.* 2015;101(3):598–605.
11. Razak F, Corsi DJ, Subramanian SV. Change in the body mass index distribution for women: analysis of surveys from 37 low- and middle-income countries. *PLoS Med.* 2013;10(1):e1001367. <https://doi.org/10.1371/journal.pmed.1001367>.
12. Vaezghasemi M, Razak F, Ng N, Subramanian S. Inter-individual inequality in BMI: an analysis of Indonesian Family Life Surveys (1993–2007). *SSM-Popul Health.* 2016;2:876–88.

13. Corsi DJ, Neuman M, Finlay JE, Subramanian S. Demographic and health surveys: a profile. *Int J Epidemiol*. 2012;41(6):1602–13.
14. MEASUREDHS. DHS Overview. <http://www.measuredhs.com/What-We-Do/Survey-Types/DHS.cfm>. Accessed 3 Dec 2017.
15. Rutstein SO. The DHS wealth index: approaches for rural and urban areas. Washington, DC: Macro International Inc; 2008.
16. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods*. 2012;17(2):228.
17. Ball K, Crawford D. Socioeconomic status and weight change in adults: a review. *Soc Sci Med*. 2005;60(9):1987–2010.
18. Hoffman L. Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivar Behav Res*. 2007;42(4):609–29.
19. Vallejo G, Fernández P, Cuesta M, Livacic-Rojas PE. Effects of modeling the heterogeneity on inferences drawn from multilevel designs. *Multivar Behav Res*. 2015;50(1):75–90.
20. Dedrick RF, Ferron JM, Hess MR, Hogarty KY, Kromrey JD, Lang TR, et al. Multilevel modeling: a review of methodological issues and applications. *Rev Educ Res*. 2009;79(1):69–102.
21. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. Newcastle upon Tyne: Sage; 2002.
22. Lewontin R, Levins R. Schmalhausen's law. *Capital Nat Soc*. 2000;11(4):103–8.
23. Schwartz S, Diez-Roux R. Commentary: causes of incidence and causes of cases—a Durkheimian perspective on Rose. *Int J Epidemiol*. 2001;30(3):435–9.
24. Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat Med*. 1999;18(12):1467–74.
25. Ogden CL, Yanovski SZ, Carroll MD, Flegal KM. The epidemiology of obesity. *Gastroenterology*. 2007;132(6):2087–102.
26. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in diet and lifestyle and long-term weight gain in women and men. *N Engl J Med*. 2011;2011(364):2392–404.
27. Lynch JW, Kaplan GA, Salonen JT. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Soc Sci Med*. 1997;44(6):809–19.
28. Duncan C, Jones K. Using multilevel models to model heterogeneity: potential and pitfalls. *Geogr Anal*. 2000;32(4):279–305.
29. Himmelstein DU, Levins R, Woolhandler S. Beyond our means: patterns of variability of physiological traits. *Int J Health Serv*. 1990;20(1):115–24.
30. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159(9):882–90.
31. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology*. 2005;16(1):124–9.
32. Merlo J, Asplund K, Lynch J, Råstam L, Dobson A. Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organization MONICA Project. *Am J Epidemiol*. 2004;159(12):1168–79.