**Supplementary Material: Model description**

Our model was constructed by linking three latent variables (physical activity, diet, and socio-demographic status) with objective health measures expressed by biomarkers (HDL, triglycerides, total cholesterol) and anthropometric measures (BMI, waist circumference, high blood pressure status). We used family history of diabetes and consanguinity as variables of interest and used age and gender as covariates.

The first dimension measured in our hypothetical model was the effect of physical activity on glucose and lipid metabolism and anthropometry. The latent variable of physical activity was constructed using five variables: hours of weekly vigorous physical activity at (1) work, and through (2) sports; hours of weekly moderate physical activity at (3) work and through (4) sports; and hours per week of (5) walking. This latent variable replicates a similar variable built by Bardenheier et al. [28] to examine risk factors for prediabetes in adults, aged 50 and over. After building this latent variable, we modeled the effects that physical activity has on (1) HDL, (2) triglycerides, (3) total cholesterol, (4) BMI, (5) waist circumference, and (6) high blood pressure status. These effects were adapted from a study by Stringhini et al. [38] that analyzed risk factors for diabetes and their links to physical activity.

Similar to Bardenheier et al.'s study [28], our dietary latent variable was built using Health Eating Index scores. The limited scope of the STEPS survey meant that we could only incorporate (1) fruit and (2) vegetable intake into our latent variable for diet. Our dietary latent variable also incorporated self-reported numbers for the days per week the subjects ate outside of their homes (3) and the frequency of fast food consumption per week (4). We modeled the effects that diet might have on (1) HDL, (2) triglycerides, (3) total cholesterol, (4) BMI, (5) waist circumference, and (6) high blood pressure status. These effects were selected based on previous studies by Ezzatti et al. [36] and Hu [37] who analyze dietary risk factors for diabetes and their relationships to dietary patterns.

The final aspect measured in our hypothetical model was the effect of socio-demographic factors on glucose and lipid metabolism and anthropometry. This socio-demographic latent variable and its effects were constructed based on models presented by Bardenheier et al. [28] and Hu [37]. The latent variable for socio-demographic factors was built using (1) educational category, (2) marital status, and (3) smoking status. Following, we modeled the effects that socio-demographic factors have on (1) HDL, (2) triglycerides, (3) total cholesterol, (4) BMI, (5) waist circumference, (6) high blood pressure status, (7) physical activity, (8) diet, and (9) diabetic status.

After building the three latent variables, we modeled their effects on diabetic status based on previous models constructed by Bardenheier et al. [28] and Hu [37]. We modeled the effects that (1) HDL, (2) triglycerides, (3) total cholesterol, (4) BMI, (5) waist circumference, (6) high blood pressure status,

consanguinity (7) and smoking (8) have on diabetic status. We were also interested in the links between obesity and metabolic biomarkers that might be mediating effects on diabetic status. Consequently, we also modeled the effects of (1) HDL, (2) triglycerides, and (3) total cholesterol on BMI, as well as the effects of (1) HDL, (2) triglycerides, (3) total cholesterol and (4) BMI on waist circumference. Furthermore, we also modeled the effects of (1) HDL on total cholesterol, (2) HDL on triglycerides, (3) waist circumference on high blood pressure status and (4) BMI on HDL. The last effects modeled were the effects that a family history of diabetes has on (1) HDL, (2) total cholesterol, (3) high blood pressure status and (4) diabetic status. Age was used as covariate in (1) physical activity, (2) diet, (3) socio-demographic status, (4) HDL, (5) tryglicerides, (6) total cholesterol, (7) BMI, (8) waist circumference, (9) high blood pressure status, (10) diabetic status and (11) smoking status. Sex was used as covariate in (1) physical activity, (2) diet, (3) socio-demographic status, (4) HDL, (5) tryglicerides, (6) total cholesterol, (7) BMI, (8) high blood pressure status, (9) diabetic status and (10) smoking status.