



OPEN

## Phenotypic decanalization driven by social determinants could explain variance patterns for glycemia in adult urban Argentinian population

María Alejandra Petino Zappala<sup>1✉</sup>, Guillermo Folguera<sup>1</sup> & Santiago Benitez-Vieyra<sup>2</sup>

Type 2 diabetes, one of the major causes of death and disability worldwide, is characterized by problems in the homeostasis of blood glucose. Current preventive policies focus mainly on individual behaviors (diet, exercise, salt and alcohol consumption). Recent hypotheses state that the higher incidence of metabolic disease in some human populations may be related to phenotypic decanalization causing a heightened phenotypic variance in response to unusual or stressful environmental conditions, although the nature of these conditions is under debate. Our aim was to explore variability patterns of fasting blood glucose to test phenotypic decanalization as a possible explanation of heightened prevalence for type 2 diabetes in some groups and to detect variables associated with its variance using a nation-wide survey of Argentinian adult population. We found patterns of higher local variance for fasting glycemia associated with lower income and educational attainment. We detected no meaningful association of glycemia or its variability with covariates related to individual behaviors (diet, physical activity, salt or alcohol consumption). Our results were consistent with the decanalization hypothesis for fasting glycemia, which appears associated to socioeconomic disadvantage. We therefore propose changes in public policy and discuss the implications for data gathering and further analyses.

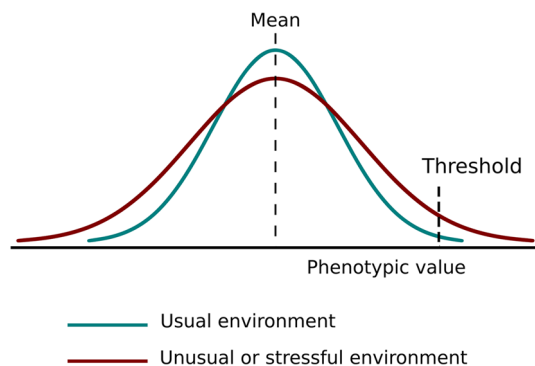
Diabetes is a health condition of global concern that afflicts an estimate of 422 million people worldwide, mostly from low-and middle-income countries<sup>1</sup>, and is predicted to become a problem affecting one-third of the world population in the next generations<sup>2</sup>. This chronic metabolic condition is characterized by high levels of blood sugar, which can cause heart disease, nerve damage, kidney failure, vision loss, problems during pregnancy, tissue damage requiring leg amputation, and generally increase the risk of disease complications and premature death<sup>1,3</sup>. In addition, in the last year, a higher susceptibility to adverse outcomes for COVID-19 has been reported for people with diabetes or high blood sugar<sup>4,5</sup>.

Type 2 diabetes, the most common kind<sup>6</sup>, is caused by insulin resistance or the relatively inadequate production of insulin for a given patient's demand. Patients usually require medication, changes in the diet, physical activity and regular checking to maintain their blood sugar levels on a healthy range.

Much work has been done to uncover and describe risk factors for type 2 diabetes or high blood sugar and most public health approaches focus on the spread of information about “healthy habits”, i.e. adequate consumption of fruits and vegetables, regular physical activity, avoidance of tobacco, alcohol and salty foods, and weight control<sup>1,3</sup>. Although socioeconomic factors related to increased prevalence of diabetes have been described, it is assumed that their effect occurs through the lack of resources (income, time and/or information) to exercise the aforementioned recommendations<sup>3</sup>.

**Decanalization.** One of the hypotheses seeking to explain complex diseases such as type 2 diabetes points to the influence of environmental stressors in phenotypic variance. This hypothesis is based on the phenomenon

<sup>1</sup>Facultad de Ciencias Exactas y Naturales, Facultad de Filosofía y Letras, Universidad de Buenos Aires, CONICET, CABA, Intendente Güiraldes 2160, Pabellón II, C1428EGA Buenos Aires, Argentina. <sup>2</sup>Instituto Multidisciplinario de Biología Vegetal (IMBIV), Universidad Nacional de Córdoba, CONICET, Córdoba, Argentina. ✉email: mapz@de.fcen.uba.ar



**Figure 1.** Representation of phenotypic (de)canalization. Expected distribution for the trait values in a hypothetical population under usual and unusual or stressful environments, determining canalized and decanalized phenotypic distributions. A heightened phenotypic variance results in a higher proportion of individuals with extreme phenotypic values at both ends of the distribution.

of canalization, a concept first coined by Waddington<sup>7</sup> to explain the adoption of discrete cell “fates”; it was later extended to groups of individuals, to describe their relative phenotypic uniformity despite genetic or minor environmental changes<sup>8</sup>. In populations living in relatively stable environments, it is expected that stabilizing selection leads to the accumulation of epistatic interactions that restrict the population’s phenotypic variance for adaptive traits near the optima for the environmental range<sup>8</sup>. This means the trait would become “canalized” around these phenotypic values, and as a consequence a majority of the individuals within the population would present similar phenotypes spanning a restricted range. However, under stressful or uncommon environmental conditions, those epistatic interactions may become perturbed; thus, more individuals are expected to present extreme phenotypic values, leading to an augmentation of phenotypic variance at the level of the population (“decanalization”), a principle first described by Schmalhausen<sup>9</sup>. Here we will focus on this particular meaning of canalization, a term usually designating a population feature which stems from the robustness of individuals’ developments in the face of environmental change.

Figure 1 summarizes the expectations under the decanalization hypothesis for complex diseases. In stable environments, important physiological variables are tightly regulated in most individuals, so that most of them present trait values near the phenotypic optimum; i.e. the phenotype is canalized, and few individuals present extreme (possibly pathological) values. In populations subject to unusual or stressful environments, phenotypic decanalization would lead, even possibly in the absence of a change in the mean value for the trait, to a higher proportion of the individuals presenting extreme phenotypic values, with possible consequences for their health and life quality.

There are two different accounts for the decanalization hypothesis that can be applied to metabolic diseases such as diabetes in human populations. Both would consider the disease prevalence, in this case type 2 diabetes, as related to the phenotypic decanalization of an otherwise canalized trait (here, the homeostatic mechanism regulating blood glucose levels) in some groups within the population. They differ, however, in terms of which are the relevant environmental factors causing this decanalization. One of the hypotheses points towards the introduction of industrialized foods and changes in motor activity (i.e. sedentarism) as driving the augmentation of phenotypic variability for blood sugar levels in human populations and, therefore, leading to a higher prevalence of type 2 diabetes<sup>10</sup>. According to this version, it follows that efforts to tackle diabetes should be focused on the promotion of healthy habits<sup>2,10</sup>. A second version of the decanalization hypothesis was formulated by Lewontin and Levins<sup>11,12</sup>. The basis of this explanation is similar, but the authors, when discussing human populations, refer to the poor, excluded and marginalized communities suffering from multiple environmental stressors which are not under their control, that depend on the historical relationship of the populations with their environment, including social relations between individuals. According to this view, inequalities directly impact on physiological processes through a multitude of mechanisms (including psychological stress derived from the position within class hierarchies). Under the stressful conditions to which these groups are subjected, mechanisms buffering the phenotype from environmental variation break down, and latent differences between individuals become manifest<sup>11,13</sup>. According to Himmelstein et al.<sup>13</sup>, homeostatic capacity generally erodes with age through the accumulation of stressors, but differences can be found when accounting for factors such as race or income, so that underprivileged people’s homeostatic mechanisms deteriorate faster; therefore these groups present greater phenotypic variability earlier. These authors, and others focusing on the social determinants of health, argue that health policies that rely on the spread of information about healthy habits to tackle diabetes are not the best way to address the problem, and they also shift the responsibility to the individuals when the causes of disease would be of social origin and therefore outside of their control<sup>12,14</sup>. This version of the hypothesis is not incompatible with other approaches like the “psychosocial stress theory”, although it provides for a possible mechanism through which stimuli of very different kinds which are perceived as stressful (or protective) may affect the susceptibility to suffer from ill health<sup>15</sup>.

It is known that physiological levels of fasting blood glucose in humans are tightly regulated; according to current guidelines the healthy range lies between 70 and 100 mg/dL. However, to our knowledge, the decanalization

hypothesis for this trait in human populations has not been systematically assessed<sup>16</sup>. Evidence from experimental studies carried out in *Drosophila melanogaster* supports the idea that the likelihood of extreme metabolic phenotypes can be augmented by dietary changes leading to a higher phenotypic variance<sup>17</sup>. There are obvious obstacles to the testing of this hypothesis in humans, although some observational studies show results consistent with it<sup>13,18</sup>.

A corollary to both decanalization hypotheses is that trying to model individual risk for diabetes would be fruitless, as would be looking for a difference in means of blood sugar levels between groups. Therefore, different statistical approaches and, ideally, surveys designed to this end would be needed to identify groups with a higher susceptibility to diabetes and to design public health strategies to counteract this problem, an objective of the utmost importance<sup>3</sup>.

**Diabetes in Argentina.** In Argentina, according to the last National Surveys on Risk Factors for Noncommunicable Diseases the prevalence of diabetes and hyperglycemia (by self-report) has been on the rise in the last two decades<sup>19</sup>. Public health campaigns at the national and regional level have focused on the promotion of healthy habits<sup>20</sup>, and although a heightened prevalence was detected in the surveys for people with lower educational attainment and public (vs. private) health insurance, no mechanism was proposed other than lack of access to or noncompliance with dietary and physical exercise guidelines<sup>19</sup>.

In contrast to previous surveys, in order to assess the possible under-diagnosis of diabetes and high blood sugar, the last National Survey carried out in 2018 incorporated biochemical measurements, including capillary fasting blood glucose. In this work, using these determinations, we tested whether data from the Argentinian urban population support the decanalization hypothesis for glycemia, in order to explore the determinants of hyperglycemia and diabetes. Moreover, we were able to explore associations with socioeconomic and individual variables that could account for the differences observed in the patterns of fasting blood glucose. This setting provides us with a basis to discuss current public preventive approaches and propose changes in health policy and future survey design and analysis.

## Materials and methods

**Data.** The raw database from the 2018 National Survey on Risk Factors for Noncommunicable Diseases, consisting on several variables from 25,208 adults from all provinces randomly selected through a probabilistic multi-stage sample design, was obtained from the INDEC webpage. All raw data are available at <https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos-2>.

Although information on diabetes diagnosis (by self report) was available, it was not used for these analyses, as it did not allow us to test the decanalization hypothesis. Therefore, we used only the determinations of fasting blood sugar; data were filtered to exclude those cases where these determinations were not performed. The relevant explanatory variables were chosen according to bibliography and availability of data on the survey. Those were age, sex, income, educational attainment for the interviewee and head of the household, density, utilities, paid working time, daily fruits and vegetables consumed, levels of alcohol and salt intake and sedentarism (see Additional Methodological note for more details).

**Generation of new variables.** We generated a categorical variable for levels of fasting blood sugar (in mg/dL) based on the raw values provided in the survey database. However, instead of the standard ranges of fasting plasma glucose used by convention to diagnose diabetes and prediabetes<sup>1</sup>, we implemented a numeric categorization, with its levels 1 to 6 (see Additional Methodological note for more details).

Age is a well described risk factor for hyperglycemia and type 2 diabetes<sup>19,21,22</sup>. To account for its effect on blood sugar levels, we calculated residuals from the linear regression between glycemia and age and created a variable grouping absolute residuals in three categories: Normal (residual < 40), Outlier ( $40 \leq \text{residual} < 80$ ), roughly representing  $Q1 + IQR$  and  $Q3 + IQR$ , and Extreme (residual  $\geq 80$ ), approximately  $Q1 + 2.5 \times IQR$  and  $Q3 + 2.5 \times IQR$ . This variable serves to group individual values that deviate from the estimate in both directions and is therefore useful to detect decanalization.

Monthly household income was also categorized in four levels: high or A (income  $\geq 40,000$  pesos argentinos), middle or B ( $20,000 \leq \text{income} < 40,000$ ), low or C ( $10,000 \leq \text{income} < 20,000$ ) and very low or D (income < 10,000). At the time the survey was performed, the cutoff for indigence for an average family (two adults, two children) in Buenos Aires was of 10,122 pesos, approximately corresponding to 280 US Dollars<sup>23</sup>. In this sense an average family in group D would be considered as indigent.

Finally, Coefficients of Variation for glycemia were calculated as  $s/\bar{x}$ , being  $s$  and  $\bar{x}$  the estimators of standard deviation and mean for each group.

**Statistical analyses.** All statistical analyses were performed in Rstudio<sup>24</sup> using R version 3.6.3<sup>25</sup>.

We applied several statistical analyses to test the decanalization and alternative hypotheses. First, we applied a “traditional” approach, i.e. we tested whether average glycemia values per individual can be predicted by the explanatory variables, which were detailed above. Contrarily to the decanalization hypothesis, which would search for an increased within-group variance to detect factors behind increased susceptibility to diabetes, all of these “traditional” analyses are aimed at modeling individual risk for diabetes.

For this approach, we applied supervised machine learning algorithms such as random forest (RF), Gradient Boosting Machine<sup>26</sup>, and regression Support Vector Machine<sup>27,28</sup>. These analyses were performed with randomForest, caret, and e1071 R packages<sup>29–31</sup>. See Additional Methodological note for more details.

Second, to test the predictions of the decanalization hypothesis and to inquire the explanatory variables influencing diabetes susceptibility, we first describe the multivariate space for our data. A Categorical Principal

Component Analysis (PCA) was performed using the `princals` function from the `Gifi` R package<sup>32,33</sup>, where all explanatory variables excepting sex were considered ordinal. No variables related to blood sugar level were used as active variables in this analysis. Therefore, the obtained components represented a multivariate space defined by the explanatory variables, where we mapped the Normal, Outlier, and Extreme residuals described above. As we used absolute residuals, those categories represent the variability of glycemia in relation to its expected value. Thus, they are an approximation to groups with different levels of decanalization, each one characterized by broader deviations from expectation. To test whether individuals from the Normal, Outlier, and Extreme groups occupied distinct sectors of the socioeconomic multivariate space, the scores for individuals in all PCA components were subjected to a Permutational Multivariate Analysis of Variance (PERMANOVA<sup>34</sup>) using the `adonis` function of the `vegan` R package<sup>35</sup>, having previously checked the assumption of multivariate homogeneity of variances (using functions `betadisper` and `permutest` from the same package). Bonferroni-adjusted pairwise comparisons were performed with the `pairwise.adonis` function<sup>36</sup>.

In the same multivariate space of socioeconomic explanatory variables, we determined regions of high and low variance for fasting blood glucose. With this aim, we used a procedure similar to a running average in two dimensions. The algorithm determines neighborhoods in the multivariate space and then computes the standard deviation for fasting blood glucose for all the observations within. This function returns a vector containing the running standard deviation for the neighborhood around each observation. This analysis was inspired by spatial analysis, in which marginalized neighborhoods should show higher values of standard deviation in glycemia. Our data sadly lacked detailed geographical information, but through PCA we could construct “neighborhoods” of individuals that share certain socioeconomic characteristics.

After identifying through the PCA and running standard deviation the variables that better explained differences in neighborhood variability for glycemia, we tested whether these variables were associated to different patterns of blood sugar levels. With this aim, we performed a log likelihood ratio test of independence to explore if the individuals with normal, outlier or extreme glucose residual values were equally distributed in the different income and educational attainment groups. using the `GTest` function from the `DescTools` package<sup>37</sup>.

**Informed consent.** All authors prepared, read and agreed with the content, gave consent to submit and obtained consent from the responsible authorities at their respective institutes.

**Ethical approval.** This research involves only the use of public anonymized data and therefore was granted exemption from ethics committee approval. The study complies with the principles of the Declaration of Helsinki.

## Results

First, data from the National Survey were filtered. From a total of 4477 individuals on which blood glucose was measured, 4115 individuals were retained for which complete records were available for all variables (Table 1, Additional File 1). Distributions for all covariates were similar for the group of urban individuals on which blood glucose was not measured (Additional File 2). Given the reported relationship between glycemia and age, which was also confirmed by our exploratory analyses, we calculated residuals from the linear regression between glycemia and age and also created a categorical variable for groups according to absolute residuals (see “Materials and methods” section).

First, we performed the “traditional” analyses to model raw values or residuals for fasting blood glucose measurements. We found no regression algorithm that could satisfyingly fit our data, as they all explained a low proportion (< 10%) of variance for both variables (Additional File 3).

8 components of the Categorical Principal Component Analysis explained 81.1% of the total variance in glycemia (loadings are presented in Additional File 4).

The first principal component was negatively correlated with educational attainment (both for the interviewee and head of the household) and income, followed by working hours, while the second component presented higher loadings for density of inhabitants in the household and interviewee’s age (Additional File 4). Variables related to the usual lifestyle recommendations (level of physical activity, diet and alcohol consumption) did not have high loadings for the first two components, although they were the most correlated with the fourth component. Therefore, the first two components were mainly determined by socioeconomic variables while lifestyle influenced the third and fourth ones. The fifth and seventh component were mostly associated with salt and alcohol consumption, respectively, with the rest of the variables showing lower loadings (Additional File 4).

Superimposing smooth density estimates for Normal, Outlier or Extreme groups for fasting blood glucose residuals over the PCA biplot allowed us to explore how they differ in their characteristics for the explanatory variables. Figure 2a shows distinct groups partially separated by the first and second Principal Components; there was a higher density of individuals with extreme residual values in the sector corresponding to a low income and low educational attainment (both for the interviewee and the head of the household). The higher density for the normal group was found in the region of higher educational attainment, higher income and lower age. No clear pattern was found regarding lifestyle variables, and Fig. 2b shows that the three groups mostly overlapped for the third and fourth Principal Components. No clear patterns were found for the rest of the components (Additional File 5). Similar results were obtained by using residuals of the linear regression between glycemia and age (not shown), when removing the individuals who reported following a medical treatment to control glucose values ( $n = 312$ ) or those already diagnosed with diabetes ( $n = 295$ ) (Additional File 6).

To analytically determine whether there were distinct regions of higher density of normal, outlier or extreme values for fasting blood glucose in the space determined by the Principal Components, we performed a PERMANOVA, after confirming no significant departures from homogeneity of multivariate dispersion in this

Covariates	Mean $\pm$ SD or %
Age (years)	50.2 $\pm$ 16.1
Sex	Male: 41.55%
	Female: 58.45%
Max. educational attainment: interviewee	Incomplete elementary school: 8.41%
	Incomplete high school: 33.90%
	Complete high school and above: 57.69%
Max. educational attainment: head of household	Incomplete elementary school: 9.28%
	Incomplete high school: 35.36%
	Complete high school and above: 55.36%
Monthly income for the household (Argentine Pesos)	23,369 $\pm$ 18,313
Working time (weekly hours)	No working time: 38.5%
	0 to 35: 22.2%
	35 to 45: 23.1%
	> 45: 16.2%
Density (inhabitants/household rooms)	1.00 $\pm$ 0.70
Utilities (from 0 = no access to gas, water or sewage system to 1 = access to gas, water and sewage system)	0.64 $\pm$ 0.17
Level of physical activity	Low: 46.2%
	Intermediate: 36.8%
	High: 17.0%
Sedentarism (daily minutes spent sitting)	261.4 $\pm$ 174.0
Daily fruit and vegetable consumption (portions)	2.03 $\pm$ 1.62
Salt consumption	Does not add salt to meals: 21.6%
	Sometimes adds salt to meals: 50.8%
	Frequently adds salt to meals: 17.4%
	Always adds salt to meals: 10.2%
Alcohol consumption	No problematic consumption: 82.9%
	Regular or episodic problematic consumption: 13.2%
	Regular and episodic problematic consumption: 3.8%

**Table 1.** Summary for all covariates used as explanatory variables in the analyses.

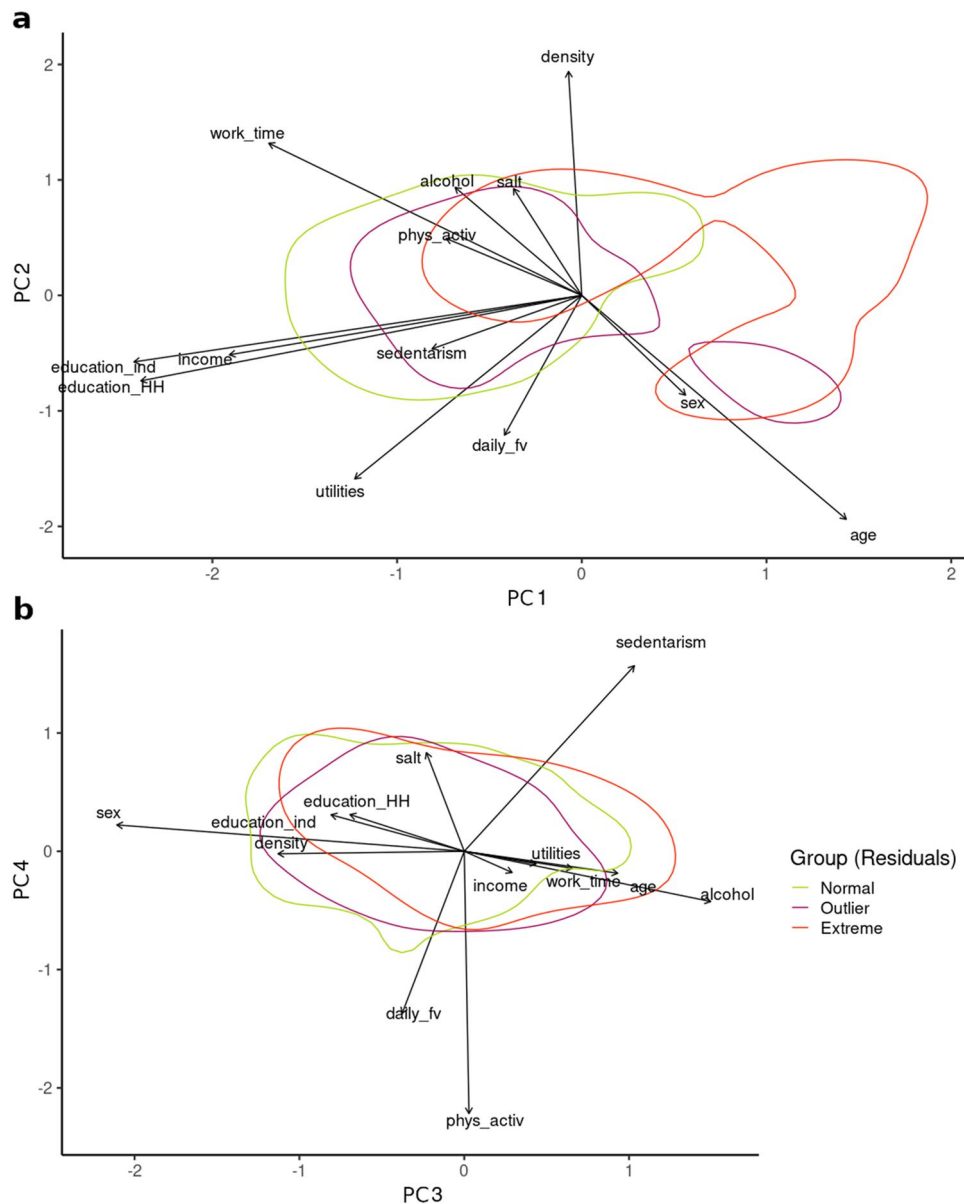
multivariate space ( $F_{2,4112} = 0.729$ ;  $p = 0.47$ ). The PERMANOVA indicated that there were significant differences between Normal, Outlier and Extreme groups of blood glucose residuals for their position in the multidimensional space regarding all components retained in the PCA ( $F_{2,4112} = 20.95$ ;  $p = 0.001$ ). Indeed, the three groups differed significantly in pairwise comparisons (Additional File 7).

The calculations of the running standard deviations showed a clear correlation between the first principal component and heterogeneity for fasting blood glucose (Fig. 3), with the regions corresponding with lower income and educational attainment presenting neighborhoods with a higher variance for the trait. For the third and fourth components, a small region of high variance was found corresponding to the higher scores for the third component (higher alcohol consumption, older age, male gender), but no clear tendency was apparent overall (Fig. 4). No defined pattern was observed either for the rest of the components (Additional File 8). These results did not change when correcting by running average (not shown).

As income and educational attainment appear as highly correlated relevant explanatory variables, we formed four income categories (see “Materials and methods” section) and estimated Coefficients of Variation of blood glycemia for all income and educational level combinations. Figure 5 shows the expected results: a trend for higher variability in low income groups. Also a reverse relationship between variability of glycemia and educational attainment can be found, although the latter is not clear in the higher and lower income groups, probably due to the scarcity of interviewees in the highest income—lowest educational attainment and lower income—higher educational attainment groups.

Finally, to analytically test if the distribution within Normal, Outlier and Extreme groups for the residuals from the linear regression between glycemia and age is independent of these socioeconomic variables, we formed groups by combining categorized income with individual educational attainment and performed a G-test for independence. We found significant deviations from independence ( $G\text{-test}_{df=22} = 68.69$ ;  $p = 1.06e-06$ ), indicating that the proportion of individuals within the Normal, Outlier or Extreme groups depends on the income-educational attainment combination. The relative contributions to the statistic for each category can be found in Fig. 6. For the high income and high education group, individuals in the Normal category for the regression between glycemia and age were more frequent than expected, a relationship that held (albeit less markedly) in the medium–low income and high education (B-3 and C-3) groups. WOn the other hand there was an excess of individuals at the Extreme and Outlier categories in the lower income and lower education groups. We could not





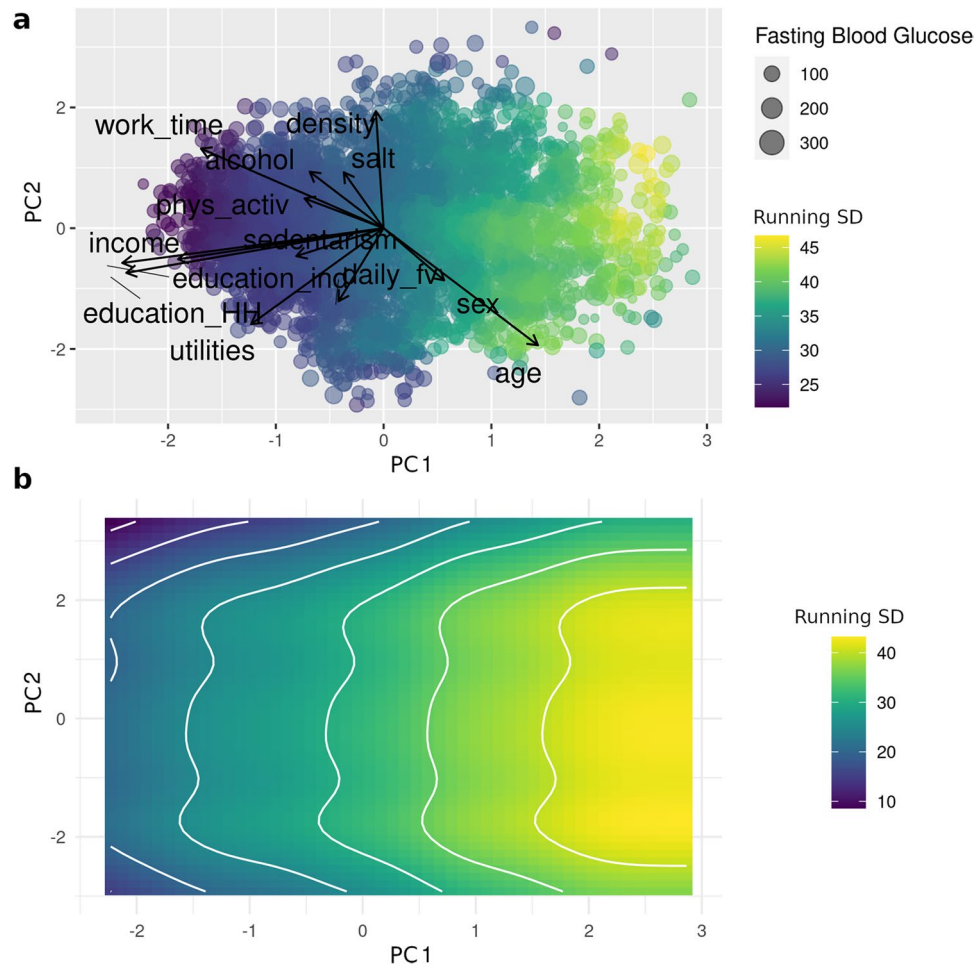
**Figure 2.** Curves for Normal, Outlier and Extreme glycemia residuals. Biplots for the first and second components (a) and third and fourth components (b) of the PCA. Density curves for Normal, Outlier and Extreme groups (bins = 2) are superimposed.

evaluate if this trend held for the high income—low educational attainment and lower income—high educational attainment categories, as they contained comparatively few individuals.

Finally, we wanted to analyze whether these statistical trends also had a clinical significance, meaning whether the distribution of blood glycemia values in the different income-education groups represented particular patterns of outcomes with impact on life quality. As expected, a higher proportion of individuals in the lower income and lower educational attainment groups presented levels of glycemia within the higher (4–6) categories (Fig. 7). It may be noted that low levels of fasting glycemia (< 70 mg/dL) are also more prevalent in the lower income groups.

## Discussion

Here we have shown how data from an Argentinian National Health Survey fit the hypothesis of phenotypic decanalization (in this case, of fasting blood glucose values) in response to socioeconomic variables, with a gradient in which the most disadvantaged groups show a greater variance than those in the least disadvantaged position as defined by income and educational attainment. Our results contrast with the dominant conception according to which socioeconomic situation acts through making the adoption of healthy habits not feasible<sup>1,3</sup>. The data here analyzed show little to no evident relationship between lifestyle variables and variability for blood



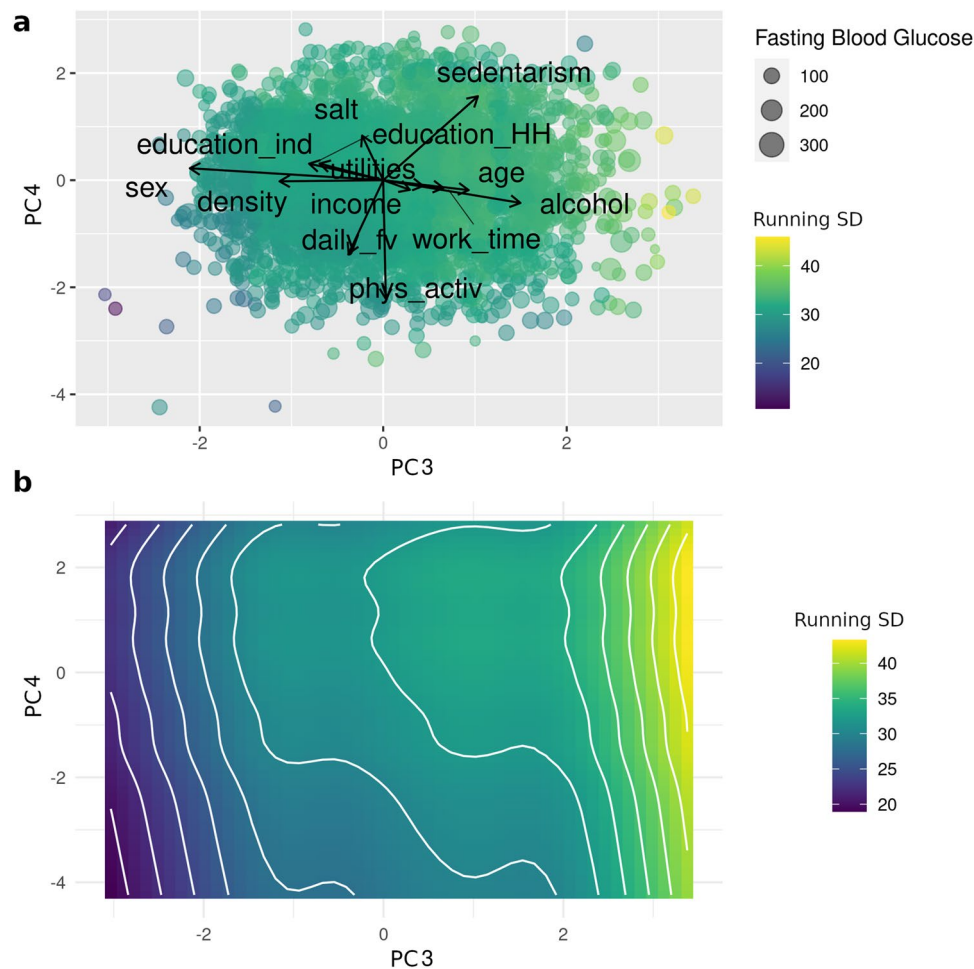
**Figure 3.** Running standard deviation for fasting glycemia in the first and second components of the PCA. **(a)** Biplot for the first and second components of the PCA; point size correlates with fasting blood glucose levels (mg/dL) and color scale indicates levels of running standard deviation (mg/dL). **(b)** GAM smoothed curves for the running standard deviation.

sugar values. It should be noted, however, that socioeconomic variables surely condition the adoption of healthy habits to some extent, as some answers to the survey indicate<sup>19</sup>.

These results are relevant for several different reasons. First, they are important in the statistical sense because of the issues they raise as a corollary: in case decanalization plays a role in disease etiology, predicting individual glycemic values would be ineffective for the affected subpopulations, as the augmentation of variance would make such predictions unreliable for individuals. However, robust patterns of variation would emerge as a group property. This is consistent with our inability to find any “traditional” model that would help explain or predict individual blood sugar values reliably. In fact, a higher group variance implies a heightened individual probability of both hyper and hypoglycemia. Therefore, if the decanalization hypothesis is true, a higher incidence of diabetes or high blood sugar could be found even without a change in means of glycemia by group, and this fact has important implications regarding statistical methods to search for factors affecting health and the design of population surveys.

But this finding is not only relevant in the statistical sense; these higher variances determine groups in which a higher proportion of individuals suffer from levels of blood glucose that are associated with pathology, or significant susceptibility to develop a health condition, as we have shown in Fig. 7. These high fasting blood sugar values lower quality of life, increase the risk of short and long-term complications, some even life-threatening<sup>1</sup>, and worsen the prognosis for several other conditions<sup>4,5,38</sup>.

There are a number of advantages of this kind of analysis. First, it overcomes some of the most frequent critiques to traditional epidemiological research<sup>39</sup> such as that it does not propose mechanisms behind increased prevalence, and therefore the real causes of ill health may remain obscure. Other objection is that traditional epidemiological analyses tend to disregard, minimize or simplify social and economic variables<sup>40</sup>. This is particularly troubling in the case of diseases affected by social or economic factors that may become masked by other variables more consistent with the idea of health as determined by individual (lifestyle or genetic) factors, therefore making the individuals responsible for their ill health, and promoting constant self-surveillance<sup>41,42</sup>. In contrast, our approach includes a possible mechanism behind the increased prevalence of complex disease



**Figure 4.** Running standard deviation for fasting glycemia in the third and fourth components of the PCA. **(a)** Biplot for the third and fourth components of the PCA; point size correlates with fasting blood glucose levels (mg/dL) and color scale indicates levels of running standard deviation (mg/dL). **(b)** GAM smoothed curves for the running standard deviation.

in vulnerable (or vulnerated<sup>43</sup>) populations that can accommodate a more holistic understanding of health and pathology processes, including the shifting social and economic relationships in the explanation of phenotypic changes<sup>40</sup>, and also would lead to concrete strategies and public policy recommendations, without focusing on individual behaviors, when enough data are available to detect reliable variability patterns.

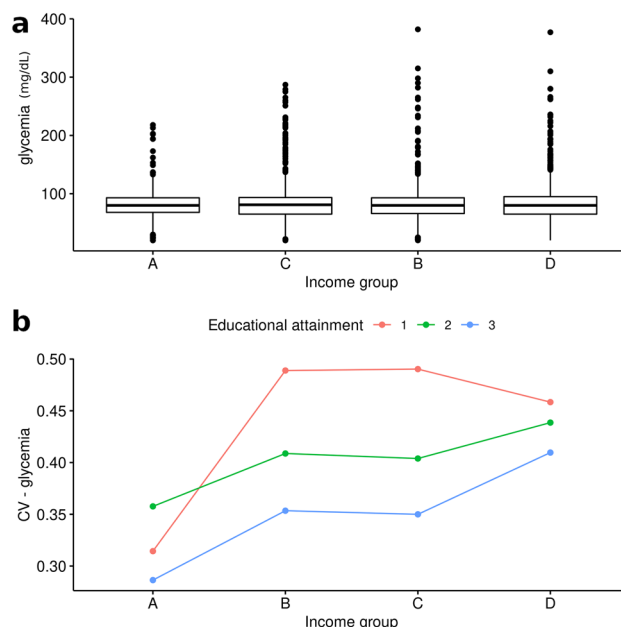
It should be noted that this approach is also not contingent on guidelines and rigid categorizations on healthy and ill individuals, which may be debatable, given the many interests and perspectives behind these discussions and the notions of health and disease themselves, and therefore result in heterogeneously adopted or changing criteria. In this sense, that this hypothesis does not rely in fixed boundaries between health and disease can be seen as an advantage and an opportunity to leave open the discussion on such boundaries.

In this sense, blood glucose determinations are an inestimable tool, as previous surveys only accounted for high blood glucose or diabetes by self-report as a categorical variable. Also, as they were performed in a probabilistic fashion, they provide a sample that is representative of the adult urban population of Argentina. Should this methodology be retained for future surveys, it will also allow for a follow-up of patterns of diabetes in the Argentinian population.

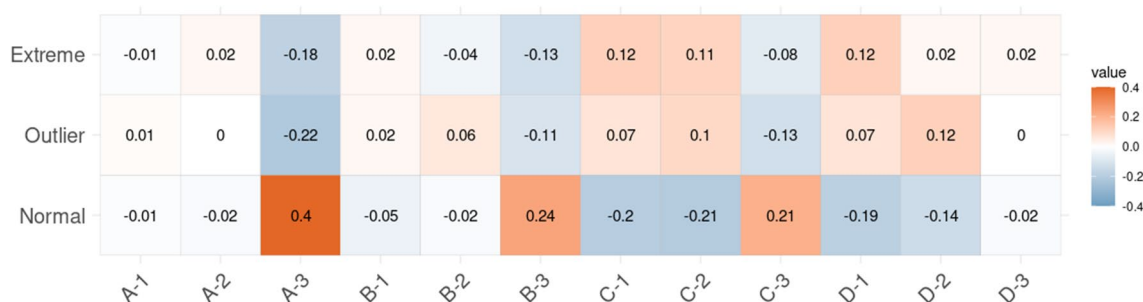
However, our analysis has some disadvantages and limitations. First of all, biochemical determinations were only carried out in interviewees from districts with a population higher than 150.000<sup>19</sup>. Therefore, our conclusions can only apply to an urban population, as previously mentioned. Another limitation on the survey concerns age, since all interviewees were adults (age  $\geq 18$  years), although this could in fact be advantageous because it weakens the weight of type 1 diabetes as a confounding factor, as it is usually detected at an earlier age (see below).

Also, the survey lacked detailed information on the geographical location of interviewees. Even as we could detect groups defined by their multivariate socioeconomic space, geographical information would have allowed us to find “natural” groups, clusters of interviewees that, we would expect, shared a more similar environment. This data would probably condense the information on general quality of life that is not available in the survey in its present form (type of neighborhood, exposure to pollutants, or other stressful environmental conditions not covered by the questionnaire).





**Figure 5.** Fasting glycemia and Coefficients of Variation by income group. Above: boxplots for values of glycemia (mg/dL) for the four income groups. Below: Coefficients of Variation for glycemia for the four income groups and the three levels of educational attainment.

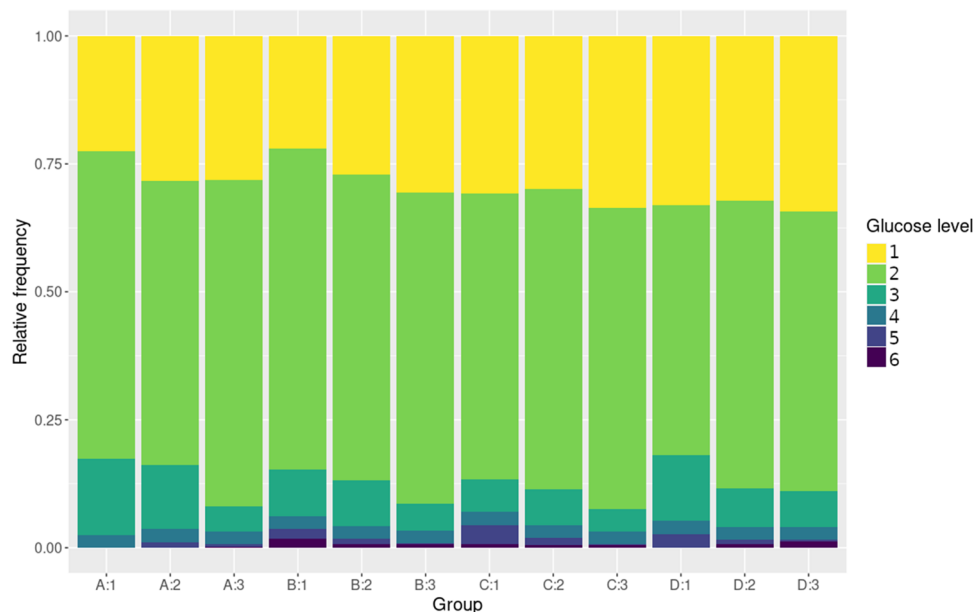


**Figure 6.** Results of the frequency analysis for categorized glycemia residuals. Residuals relative to G-statistic value for all combinations of income and educational attainment. Red color indicates that the frequency of individuals in the group is higher than expected if blood glucose residual levels were independent of income and educational groups, while blue denotes a lower frequency than expected under the assumption of independence.

Other modifications that could be suggested to better characterize the individuals' situation include gathering more detailed information on housing conditions, employment stability, unpaid workload and also subjective assessments on life quality.

Another downside of this approach is the inability to distinguish, in case a pattern of high variability of glycemia is found, if it could correspond to a subpopulation with a higher incidence of type 1 or gestational diabetes. However, as mentioned above, type 1 diabetes is usually detected in childhood or adolescence<sup>44</sup> and has a low prevalence, with an estimate for Argentinian children and adolescents of 8.6 per 100,000 people<sup>45</sup> and of 15 per 100,000 people worldwide<sup>46</sup>. Estimates of its heritability are also somewhat higher than for type 2 diabetes<sup>47</sup>. Therefore it is safe to assume that it should not affect our results to a great extent. On the other hand, since the Risk Factor Survey does not account for a possible pregnancy, a high blood sugar level caused by gestational diabetes cannot be ruled out. According to the survey, from the total of women with a self-reported history of diabetes, 21.6% were pregnant when diagnosed. However, it would probably not change our present results, as a low percentage of women would have been pregnant at the time of the survey, when measurements were taken. In fact, it has been estimated that type 2 diabetes represents up to 95% of the cases in developing countries<sup>6</sup>, and therefore it is expected to account for the majority of the cases recorded in the survey. It should be noted, moreover, that removing individuals with a confirmed diabetes diagnosis, regardless of the type, did not alter the general variability patterns observed here, so we could expect this method to be useful to detect regions of high variability in the general population.

Finally, our design does not allow us to determine causality, although the similar results seen by other researchers in different countries are suggestive that the effect of these particular factors in the distribution of



**Figure 7.** Levels of fasting glycemia per income and educational attainment. Relative frequency of individuals presenting each glucose level at each income and educational attainment combination.

health outcomes can be pervasive in different groups<sup>14,48</sup>. In particular, the patterns observed for the different categories of educational attainment and family income (Figs. 5, 6) are worth mentioning. Because of our setting, we cannot differentiate between two possibilities: (i) that lower educational attainment is indicative of a situation of long-term vulneration so that the relatively higher prevalence of outlier/extreme values we have seen in middle-income individuals with lower educational attainment compared to those with higher education relates to this long-term effects; and/or (ii) that higher educational attainment acts as a “protective” factor (for example, through increased social bonding, favoring autonomy and empowerment, and/or securing stable living conditions). Indeed, other researchers have already proposed that access to higher education may counterbalance to some extent the heightened susceptibility to ill health due to low income<sup>14</sup>. Our results are consistent with both explanations, although the pattern is less clear in the lower-income group, possibly because of the low number of individuals in the high educational attainment category, or else because even higher education cannot offset the stresses related to a family income below indigence threshold.

It should be noted, also, that even if income, and to some extent educational attainment, are recognized widely as affecting health in different contexts, our approach does not search for universal laws governing health. However, we argue that this methodology can be useful to detect local patterns of health outcomes, and that these insights should be integrated into a pluralistic framework including research at different levels, combining these more general quantitative analyses with in-depth, qualitative studies, to better understand the processes behind health disparities and how these are experienced by individuals and communities in different contexts<sup>49</sup>.

Although lower socioeconomic status has been associated with poorer glycemic control in people diagnosed with diabetes<sup>50</sup>, to our knowledge, this is the first work that addresses a possible relationship between these socioeconomic variables and the decanalization hypothesis for fasting blood glucose in the general population (i.e. as part of the etiological mechanism). It is worth noting that, in case these results hold in further studies, the conclusions reached regarding public policy would be very similar to the ones proposed by Marmot<sup>14</sup>: adequate living standards in the broad sense should be granted, which means reinforcing support to vulnerated groups. If the stresses related to lack of stability (of housing, working conditions, etc.) and a subordinate social position are responsible for the heightened incidence of diabetes, policies that focus on healthy habits and/or access to health care will not suffice to prevent further cases. Himmelstein et al.<sup>13</sup>, who found results consistent with the decanalization hypothesis for blood pressure, suggest reducing the exposure to these stressors and protecting at-risk populations from environmental fluctuations. According to them, the strengthening of social networks and the access to behavioral “buffering” pathways would partly unload the burden posed on the physiological subsystem. Marmot<sup>14</sup> similarly argues that empowerment acts as a protecting factor and suggests policies to improve income, education, housing stability, working conditions, etc., which may have more meaningful impacts than current health policies for diseases like type 2 diabetes, a growing economic burden both for individuals and health systems<sup>51</sup>.

We believe the decanalization hypothesis is a promising model for the study of the complex and historically changing patterns of susceptibility to ill health, particularly to chronic multi-factorial ailments related to homeostasis, which can incorporate biological (ecological, developmental, evolutionary) and socioeconomic determinants. We believe our approach, however, can be improved. Changes in data gathering methodology may benefit from interdisciplinary work to incorporate useful variables to further test this hypothesis, as current survey designs and statistical methods are not well-tailored to this end. However, this is not the only reason

why we want to highlight the importance of the dialogue between social, biological, and health sciences. The limitations generated by disciplinary fragmentation on epidemiology have already been described, and different counter-strategies proposed<sup>12,39,40,49</sup>. The notion of human populations as immersed in their socio-ecosystems, constrained by social and material conditions that are affected by a myriad of factors with a particular history and contexts, is necessary to understand the processes behind health and disease through the incorporation of different approaches and methodologies. This requires avoiding the social-biological dichotomy and overcoming disciplinary fragmentation. This issue has been raised for several subdisciplines of biology<sup>52</sup> including those involved in human health<sup>12,40,53,54</sup>. But moreover, our own notions and concepts of health and disease and the frameworks used to understand the complexity of the processes behind them also require problematizing and rethinking, and interdisciplinary work may fuel constant revisions of our approaches and substantiate a more pluralistic take on these issues<sup>49</sup>.

## Data availability

The datasets used in this study are referenced in the “Materials and methods” section (raw data) and uploaded as Supplementary Material.

Received: 24 June 2021; Accepted: 16 June 2022

Published online: 27 June 2022

## References

1. World Health Organization. *Diabetes* (2020). <https://www.who.int/health-topics/diabetes> (Accessed December 2020).
2. Gibson, G. *It Takes a Genome: How a Clash Between Our Genes and Modern Life is Making Us Sick* 1st edn. (FT Press, 2008).
3. World Health Organization. *Global Report on Diabetes* (2016). <https://www.who.int/publications/i/item/9789241565257> (Accessed May 2021).
4. Carrasco-Sánchez, F. J. *et al.* Admission hyperglycaemia as a predictor of mortality in patients hospitalized with COVID-19 regardless of diabetes status: Data from the Spanish SEMI-COVID-19 Registry. *Ann. Med.* **53**(1), 103–116. <https://doi.org/10.1080/07853890.2020.1836566> (2021).
5. Singh, A. K., Gupta, R., Ghosh, A. & Misra, A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. *Diabetes Metab. Syndr. Clin. Res. Rev.* **14**(4), 303–310. <https://doi.org/10.1016/j.dsx.2020.04.004> (2020).
6. World Health Organization. *Prevention of Diabetes Mellitus. Technical Reports* (WHO, 2014).
7. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**(3811), 563–565 (1942).
8. Flatt, T. The evolutionary genetics of canalization. *Q. Rev. Biol.* **80**(3), 287–316. <https://doi.org/10.1086/432265> (2005).
9. Schmalhausen, I. I. *Factors of Evolution: The Theory of Stabilizing Selection* (The Blakiston Co, 1949).
10. Gibson, G. Decanalization and the origin of complex disease. *Nat. Rev. Genet.* **10**(2), 134–140. <https://doi.org/10.1038/nrg2502> (2009).
11. Lewontin, R. & Levins, R. Schmalhausen’s law. *Capital. Nat. Social.* **11**(4), 103–108. <https://doi.org/10.1080/10455750009358943> (2000).
12. Lewontin, R. & Levins, R. *Biology Under the Influence: Dialectical Essays on Ecology, Agriculture, and Health* (Monthly Review Press, 2007).
13. Himmelstein, D. U., Levins, R. & Woolhandler, S. Beyond our means: Patterns of variability of physiological traits. *Int. J. Health Serv.* **20**(1), 115–124. <https://doi.org/10.2190/BKDL-N7DB-BDW8-DPYY> (1990).
14. Marmot, M. *The Health Gap: The Challenge of an Unequal World* (Bloomsbury, 2015).
15. Cassel, J. Psychosocial processes and “stress”: Theoretical formulation. *Int. J. Health Serv.* **4**(3), 471–482. <https://doi.org/10.2190/WF7X-Y1L0-BFKH-9QU2> (1974).
16. Gibson, G. & Lacek, K. A. Canalization and robustness in human genetics and disease. *Ann. Rev. Genet.* **54**, 189–211. <https://doi.org/10.1146/annurev-genet-022020-022327> (2020).
17. Reed, L. K. *et al.* Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila melanogaster*. *Genetics* **185**(3), 1009–1019. <https://doi.org/10.1534/genetics.109.113571> (2010).
18. Kodama, K. *et al.* Ethnic differences in the relationship between insulin sensitivity and insulin response: A systematic review and meta-analysis. *Diabetes Care* **36**(6), 1789–1796. <https://doi.org/10.2337/dc12-1235> (2013).
19. Instituto Nacional de Estadística y Censos. *4º Encuesta Nacional de Factores de Riesgo. Resultados Definitivos* (2019). [https://www.indec.gov.ar/ftp/cuadros/publicaciones/enfr\\_2018\\_resultados\\_definitivos.pdf](https://www.indec.gov.ar/ftp/cuadros/publicaciones/enfr_2018_resultados_definitivos.pdf) (Accessed May 2021).
20. Ministerio de Salud de la Nación Argentina. *Materiales Para la Población* (2021). <https://bancos.salud.gov.ar/bancos/materiales-para-poblacion> (Accessed May 2021).
21. Boyns, D. R., Crossley, J. N., Abrams, M. E., Jarrett, R. J. & Keen, H. Oral glucose tolerance and related factors in a normal population sample—I. Blood sugar, plasma insulin, glyceride, and cholesterol measurements and the effects of age and sex. *Br. Med. J.* **1**(5644), 595–598. <https://doi.org/10.1136/bmj.1.5644.595> (1969).
22. Fletcher, B., Gulanick, M. & Lamendola, C. Risk factors for type 2 diabetes mellitus. *J. Cardiovasc. Nurs.* **16**(2), 17–23 (2002).
23. Instituto Nacional de Estadística y Censos. *Valorización Mensual de la Canasta Básica Alimentaria y de la Canasta Básica Total. Gran Buenos Aires* (2018). [https://www.indec.gov.ar/uploads/informesdeprensa/canasta\\_12\\_18.pdf](https://www.indec.gov.ar/uploads/informesdeprensa/canasta_12_18.pdf) (Accessed May 2021).
24. Allaire, J. *RStudio: Integrated Development Environment for R*, Vol. 770, 394 (2012). <http://www.rstudio.com/> (Accessed May 2021).
25. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2020). <https://www.R-project.org/> (Accessed May 2021).
26. Breiman, L., Friedman, J. H., Olshen, R. & Stone, C. *Classification and Regression Trees* (Wadsworth & Brooks, 2003).
27. Vapnik, V. N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**(5), 988–999. <https://doi.org/10.1109/72.788640> (1999).
28. Noble, W. S. What is a support vector machine? *Nat. Biotechnol.* **24**(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565> (2006).
29. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**(3), 18–22 (2002).
30. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**(5), 1–26 (2008).
31. Meyer, D. *et al.* Package ‘e1071’. R J. (2019).
32. Mair, P. & de Leeuw, J. *A General Framework for Multivariate Analysis with Optimal Scaling: The R Package Aspect* (Department of Statistics, UCLA, 2009).
33. De Leeuw, J., Mair, P. & Groenen, P. J. F. *Multivariate Analysis with Optimal Scaling* (2017). [https://bookdown.org/jandeleeuw6/gif/\\_main.pdf](https://bookdown.org/jandeleeuw6/gif/_main.pdf) (Accessed May 2021).
34. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x> (2001).

35. Oksanen, J. *et al.* Package ‘vegan’. *Community Ecology Package, Versión*, Vol. 2, 1–295. (2013). <https://CRAN.R-project.org/package=vegan> (Accessed May 2021).
36. Martínez Arbizu, P. pairwiseAdonis: Pairwise multilevel comparison using adonis. *R Package Version 0.4. GitHub* (2019). <https://github.com/pmartinezarbizu/pairwiseAdonis> (Accessed May 2021).
37. Signorell, A. Package ‘DescTools’. *Version 0.99.32* (2020). <https://CRAN.R-project.org/package=DescTools> (Accessed May 2021).
38. Deshpande, A. D., Harris-Hayes, M. & Schootman, M. Epidemiology of diabetes and diabetes-related complications. *Phys. Ther.* **88**(11), 1254–1264. <https://doi.org/10.2522/ptj.20080020> (2008).
39. Susser, E. Eco-epidemiology: Thinking outside the black box. *Epidemiology* **15**(5), 519–520. <https://doi.org/10.1097/01.ede.0000135911.42282.b4> (2004).
40. La, B. J. determinación social de la salud como herramienta de transformación hacia una nueva salud pública (salud colectiva). *Rev. Fac. Nac. Salud Pub.* **31**, 13–27 (2013).
41. Broom, D. & Whittaker, A. Controlling diabetes, controlling diabetics: Moral language in the management of diabetes type 2. *Soc. Sci. Med.* **58**(11), 2371–2382. <https://doi.org/10.1016/j.socscimed.2003.09.002> (2004).
42. Rail, G. & Jette, S. Reflections on biopedagogies and/of public health: On bio-others, rescue missions, and social justice. *Cult. Stud. Crit. Methodol.* **15**(5), 327–336. <https://doi.org/10.1177/1532708615611703> (2015).
43. Cunha, T. & Garrafa, V. Vulnerability a key principle for global bioethics? *Camb. Q. Healthcare Ethics* **25**, 197. <https://doi.org/10.1017/S096318011500050X> (2016).
44. Chiang, J. L., Kirkman, M. S., Laffel, L. M. & Peters, A. L. Type 1 diabetes through the life span: A position statement of the American Diabetes Association. *Diabetes Care* **37**(7), 2034–2054. <https://doi.org/10.2337/dc14-1140> (2014).
45. Patterson, C. C. *et al.* Worldwide estimates of incidence, prevalence and mortality of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Diabetes Atlas. *Diabetes Res. Clin. Pract.* **157**, 107842. <https://doi.org/10.1016/j.diabres.2019.107842> (2019).
46. Mobasser, M. *et al.* Prevalence and incidence of type 1 diabetes in the world: A systematic review and meta-analysis. *Health Promot. Perspect.* **10**(2), 98 (2020).
47. Borchers, A. T., Uibo, R. & Gershwin, M. E. The geoepidemiology of type 1 diabetes. *Autoimmunity Rev.* **9**(5), A355–A365. <https://doi.org/10.1016/j.autrev.2009.12.003> (2010).
48. Kaplan, G. A. & Keil, J. E. Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation* **88**(4), 1973–1998. <https://doi.org/10.1161/01.CIR.88.4.1973> (1993).
49. Almeida Filho, N. *Epidemiología sin Números* (Organización Panamericana de la Salud, 1992).
50. Wysham, C. *et al.* OR14-3 The association of glycemic control and glycemic variability with treatment patterns and diabetes-related events in the adult patients (Pts) with type 1 diabetes (T1D): Results from the T1D Exchange Clinic Registry. *J. Endocr. Soc.* **3**, 14. <https://doi.org/10.1210/abstract-2019-OR14-3> (2019).
51. Bommer, C. *et al.* Global economic burden of diabetes in adults: Projections from 2015 to 2030. *Diabetes Care* **41**(5), 963–970. <https://doi.org/10.2337/dc17-1962> (2018).
52. Alberti, M. *et al.* Integrating humans into ecology: Opportunities and challenges for studying urban ecosystems. *Bioscience* **53**(12), 1169–1179. [https://doi.org/10.1641/0006-3568\(2003\)053\[1169:IHIEOA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[1169:IHIEOA]2.0.CO;2) (2003).
53. Schölermerich, V. L. & Kawachi, I. Translating the socio-ecological perspective into multilevel interventions: Gaps between theory and practice. *Health Educ. Behav.* **43**(1), 17–20. <https://doi.org/10.1177/1090198115605309> (2016).
54. Levins, R. & Dunn, M. L. Whose scientific method? Scientific methods for a complex world. *New Solut.* **13**(3), 261–274. <https://doi.org/10.2190/q4tn-q9u2-er56-3t1r> (2003).

## Acknowledgements

M.A.P.Z. is recipient of a postdoctoral scholarship from CONICET (Argentina), G.F. and S.B.V. are members of Carrera del Investigador Científico de CONICET (Argentina). We dedicate this work to the memory of Richard Lewontin and Richard Levins and thank them for being a constant source of inspiration.

## Author contributions

M.A.P.Z. and S.B.V. performed the data analyses. M.A.P.Z. prepared the figures. All authors prepared, reviewed and agreed with the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15041-9>.

**Correspondence** and requests for materials should be addressed to M.A.P.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022