

Testing the phenotypic decanalization hypothesis: social determinants of hyperglycemia and type 2 diabetes in adult urban Argentinian population

María Alejandra Petino Zappala^{1*}, Guillermo Folguera¹, Santiago Benitez-Vieyra²

¹ Facultad de Ciencias Exactas y Naturales, Facultad de Filosofía y Letras, Universidad de Buenos Aires, CONICET

² Instituto Multidisciplinario de Biología Vegetal (IMBIV), Universidad Nacional de Córdoba, CONICET

8

* Corresponding author: mapz@ege.fcen.uba.ar

10 Postal address: Intendente Güiraldes 2160, Pabellón II. Zipcode: C1428EGA, CABA, Argentina.

Phone number: 00 54 (011) 5285-8623

12 <https://orcid.org/0000-0003-3168-5025>

14

Abstract

16

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.

30

32

Introduction

34

Diabetes is a health condition of global concern that afflicts an estimate of 422 million people worldwide, mostly from low-and middle-income countries [1], and is predicted to become a problem affecting one-third of the world population in the next generations [2]. 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 [1, 3]. 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 [4, 5].

44 Type 2 diabetes, the most common kind [6], is caused by insulin resistance or a lack of production of this hormone. 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 [1, 3]. 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 [3].

54 Decanalization

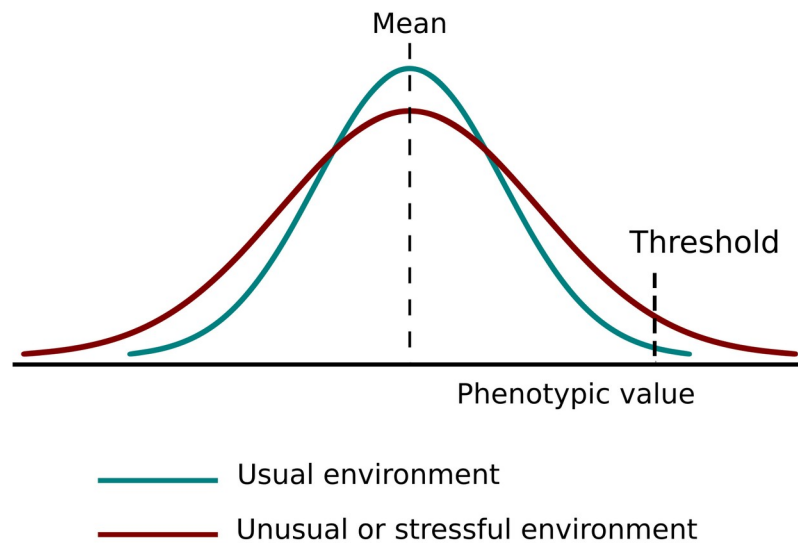
One of the hypotheses seeking to explain complex diseases such as type 2 diabetes points to the influence of environmental stressors in phenotypic variance. This hypothesis is based on the phenomenon of canalization, a concept first coined by Waddington [7] to explain the adoption of discrete and fixed "fates" by cell populations; it was later extended to groups of individuals, to describe their relative phenotypic uniformity despite genetic or minor environmental changes (in this sense, canalization is said to comprise both genetic and environmental mechanisms [8]). In populations living in relatively stable environments, it is expected that stabilizing selection leads to the accumulation of

62 epistatic interactions that restrict the population’s phenotypic variance for adaptive traits near the
optima for the environmental range [8]. This reduction of variance means the trait would become
64 “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
66 or uncommon environmental conditions, those epistatic interactions may become perturbed; thus, more
individuals are expected present extreme phenotypic values, leading to an augmentation of phenotypic
68 variance at the level of the population (“decanalization”), a principle first described by Schmalhausen
[9]. Here we will focus on this particular meaning of canalization, a term usually designing a
70 population feature which stems from the robustness of individuals’ developments in the face of
environmental change.

72

Figure 1 summarizes the expectations under the decanalization hypothesis for complex diseases. In
74 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
76 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
78 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.

80



82 **Fig. 1** Representation of phenotypic (de)canalization.

84 Expected distribution for the trait values in a hypothetical population under usual and unusual or stressful environments,
84 determining canalized and decanalized phenotypic distributions. A heightened phenotypic variance results in a higher
86 proportion of individuals with extreme phenotypic values at both ends of the distribution.

86

88 There are two different accounts for the decanalization hypothesis that can be applied to metabolic
88 diseases such as diabetes in human populations. Both would consider the disease prevalence, in this
88 case type 2 diabetes, as related to the phenotypic decanalization of an otherwise canalized trait (here,
90 the homeostatic mechanism regulating blood glucose levels) in some groups within the population.
90 They differ, however, in terms of which are the relevant environmental factors causing this
92 decanalization. One of the hypotheses points towards the introduction of industrialized foods and
92 changes in motor activity (i.e. sedentarism) as driving the augmentation of phenotypic variability for
94 blood sugar levels in human populations and, therefore, leading to a higher prevalence of type 2
94 diabetes [10]. According to this version, it follows that efforts to tackle diabetes should be focused on
96 the promotion of healthy habits [2, 10]. A second version of the decanalization hypothesis was
96 formulated by Lewontin and Levins [11, 12]. The basis of this explanation is similar, but the authors,
98 when discussing human populations, refer to the poor, excluded and marginalized communities
98 suffering from multiple environmental stressors which are not under their control, that depend on the

100 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
102 multitude of mechanisms (including psychological stress derived from the position within class
hierarchies) affecting health as a whole. Under the stressful conditions to which these groups are
104 subjected, mechanisms buffering the phenotype from environmental variation break down, and latent
differences between individuals become manifest [11, 13]. According to Himmelstein et al. [13],
106 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
108 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
110 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
112 causes of disease would be of social origin and therefore outside of their control [12, 14]. This version
of the hypothesis is not incompatible with other approaches like the “psychosocial stress theory”,
114 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 [15].

116

It is known that physiological levels of fasting blood glucose in humans are tightly regulated; according
118 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 [16].
120 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
122 phenotypic variance [17]. There are obvious obstacles to the testing of this hypothesis in humans,
although some observational studies show results consistent with it [13, 18].

124

A corollary to both decanalization hypotheses is that trying to model individual risk for diabetes would
126 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
128 identify groups with a higher susceptibility to diabetes and to design public health strategies to
counteract this problem, an objective of the utmost importance [3].

130

Diabetes in Argentina

132 In Argentina, according to the last National Surveys on Risk Factors for Noncommunicable Diseases
134 the prevalence of diabetes and hyperglycemia (by self-report) has been on the rise in the last two
136 decades [19]. Public health campaigns at the national and regional level have focused on the promotion
138 of healthy habits [20], 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
[19].

140 In contrast to previous surveys, in order to assess the possible under-diagnosis of diabetes and high
142 blood sugar, the last National Survey carried out in 2018 incorporated biochemical measurements,
144 including capillary fasting blood glucose. In this work, using these determinations, we tested whether
146 data from the Argentinian urban population support the decanalization hypothesis for glycemia, in
148 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.

150

Materials and methods

Data

154 The raw database from the 2018 National Survey on Risk Factors for Noncommunicable Diseases,
156 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
158 available at <https://www.indec.gov.ar/indec/web/Institucional-Indec-BasesDeDatos-2>.

160 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
162 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
164 bibliography and availability of data on the survey. Those were age, sex, income, educational
attainment for the interviewee and head of the household (grouped in three levels; 1 = elementary
166 school not completed, 2 = completed elementary school, high school not completed, 3 = completed
high school), density (ratio of inhabitants / number of rooms on the household), utilities (a joint
168 variable combining the information for water, gas and sewage system availability, rescaled between 0
and 1), paid working time (grouped in three levels, besides zero for unemployed or retired
170 interviewees, and not including unpaid work; 1 = time < 35 weekly hours, 2 = 35 < time < 45 weekly
hours, 3 = time > 45 weekly hours), average number of daily fruits and vegetables consumed, levels of
172 alcohol (0 = no problematic consumption, 1 = problematic regular consumption or problematic
episodic consumption, 2 = problematic regular consumption and problematic episodic consumption)
174 and salt intake (0 = does not use salt, to 3 = always adds salt to meals), levels of physical activity (1 =
low, 2 = intermediate, 3 = high) and sedentarism (daily minutes spent sitting).

176

Generation of new variables

178 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
180 glucose used by convention to diagnose diabetes and prediabetes [1], we implemented a more gradual
categorization, with its levels being hypoglycemia (glycemia<70), normal (70<=glycemia<110),
182 borderline (110<=glycemia<140), prediabetes (140<=glycemia<200), diabetes (200<=glycemia<240)
and severe diabetes (glycemia>=240).

184 Age is a well described risk factor for hyperglycemia and type 2 diabetes [19, 21, 22]. To account for
its effect on blood sugar levels, we calculated residuals from the linear regression between glycemia
186 and age , and created a variable grouping absolute residuals in three categories: Normal (residual<40),
Outlier (40<=residual<80), roughly representing $Q1 + IQR$ and $Q3 + IQR$, and Extreme
188 (residual>=80), approximately $Q1 + 2.5*IQR$ and $Q3 + 2.5*IQR$. This variable serves to group

individual values that deviate from the estimate in both directions and is therefore useful to detect
190 decanalization.

192 Monthly household income was also categorized in four levels: high or A (income \geq 40,000 pesos
argentinos), middle or B (20,000 \leq income $<$ 40,000), low or C (10,000 \leq income $<$ 20,000) and very low
194 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
196 280 US Dollars [23]. In this sense an average family in group D would be considered as indigent.

198 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.

200

202 Statistical analyses

All statistical analyses were performed in Rstudio [24] using R version 3.6.3 [25].

204 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
206 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
208 increased susceptibility to diabetes, all of these “traditional” analyses are aimed at modeling individual
risk for diabetes.

210 For this approach, we applied supervised machine learning algorithms such as random forest (RF),
Gradient Boosting Machine (GBM), and regression Support Vector Machine (SVM). These analyses
212 were performed with randomForest, caret, and e1071 R packages [26-28]. RF works by repeatedly
subsampling from the data, constructing a decision tree for each subsample (in the case of regression,
214 the response variable is discretized) and then averaging or combining the resulting trees. GBM also
works with trees, by producing an ensemble of “weak” ones into a “strong” classifier, where each
216 decision tree has a different influence according to performance [29]. SVM works by finding a
hyperplane defined by the optimal parameters to explain the response variable (i.e. minimizing the
218 distance to the maximum number of points) [30, 31].

220 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
222 Principal Component Analysis (PCA) was performed using the princals function from the Gifi R
package [32], where all explanatory variables excepting sex were considered ordinal. This analysis
224 combines linear multivariate analysis with optimal transformation of the categorical variables using
alternating least squares [33]. No variables related to blood sugar level were used as active variables in
226 this analysis. Therefore, the obtained components represented a multivariate space defined by the
explanatory variables. Then, on this multivariate space, we mapped the categorical variable that groups
228 absolute residuals from the regression between glycemia and age in three categories: Normal, Outlier,
and Extreme, as described above. As we used absolute residuals, those categories represent the
230 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
232 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
234 a Permutational Multivariate Analysis of Variance (PERMANOVA [34]) using the adonis function of
the vegan R package [35], having previously checked the assumption of multivariate homogeneity of
236 variances (using functions betadisper and permutest from the same package). Bonferroni-adjusted
pairwise comparisons were performed with the pairwise.adonis function [36].

238
In the same multivariate space of socioeconomic explanatory variables, we determined regions of high
240 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
242 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
244 observation. This analysis was inspired by spatial analysis: imagine a city where diabetes susceptibility
varies among real neighborhoods. Then, following the decanalization hypothesis from Lewontin and
246 Levins [11, 12], marginalized neighborhoods should show higher values of standard deviation in
glycemia. Our data sadly lacked detailed geographical information, but through PCA we could
248 construct “neighborhoods” of individuals that share certain socioeconomic characteristics.

250 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
252 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
254 equally distributed in the different income and educational attainment groups. using the GTest function
from the DescTools package [37].

256

258 **Results**

First, data from the National Survey were filtered. From a total of 4,477 individuals on which blood
260 glucose was measured, 4,115 individuals were retained for which complete records were available for
all variables (Table 1, Additional File 1). Given the reported relationship between glycemia and age,
262 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
264 absolute residuals (see Materials and methods).

266

Covariates	Mean ± SD or %
Age (years)	50.2 ± 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 ± 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 ± 0.70
Utilities (from 0 = no access to gas, water or sewage system to 1 = access to gas, water and sewage system)	0.64 ± 0.17
Level of physical activity	Low: 46.2%
	Intermediate: 36.8%
	High: 17.0%
Sedentarism (daily minutes spent sitting)	261.4 ± 174.0
Daily fruit & vegetable consumption (portions)	2.03 ± 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

268

First, we performed the “traditional” analyses to model raw values or residuals for fasting blood
 270 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 2).

272

We performed a Categorical Principal Component Analysis and generated a multivariate space of the
 274 explanatory individual and socioeconomic variables; we retained 8 components explained 81.1% of the

total variance. Loadings for the first 4 components (with a cutoff of 0.1) are shown in Table 2 (for all 8
276 components retained, see Additional file 3).

278

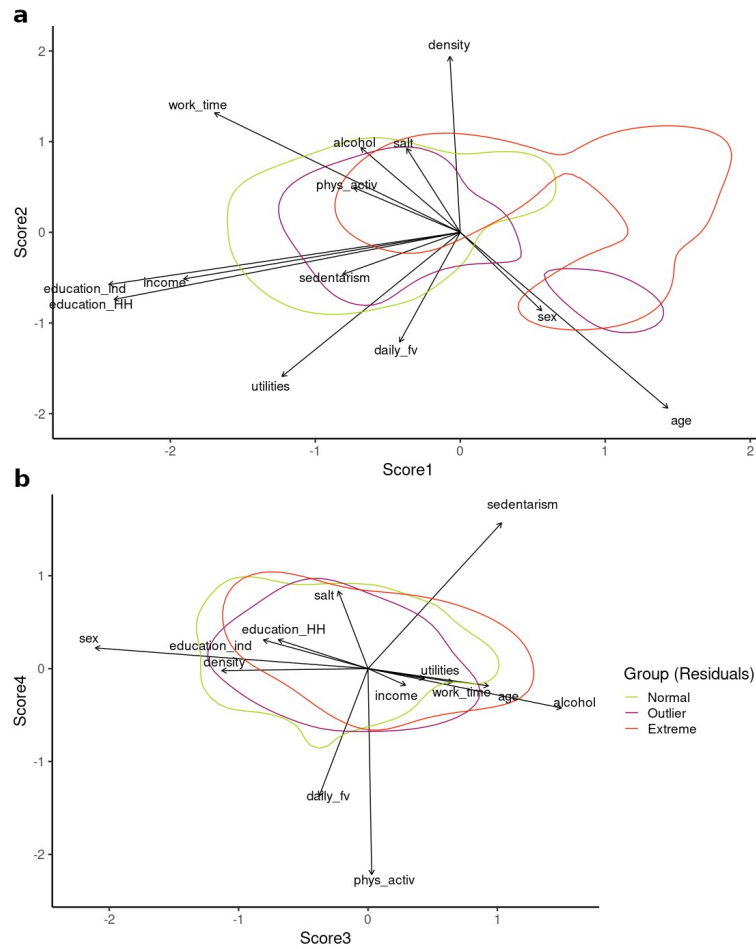
	PC1 (20.62%)	PC2 (15.12%)	PC3 (10.20%)	PC4 (8.97%)
Age	0.450	-0.665	0.301	
Sex	0.180	-0.283	-0.695	0.130
Education: interviewee	-0.814	-0.153	-0.275	0.131
Education: Head Household	-0.805	-0.208	-0.234	0.129
Income	-0.645	-0.170		
Working time	-0.547	0.460	0.173	
Density (inhabitants/rooms)		0.654	-0.375	
Utilities	-0.426	-0.519	0.145	
Physical activity	-0.238	0.178		-0.724
Sedentarism	-0.292	-0.196	0.373	0.481
Daily fruit & veg consumption	-0.144	-0.407	-0.186	-0.476
Salt consumption	-0.151	0.313		0.279
Alcohol consumption	-0.225	0.324	0.496	-0.171

Table 2. Loadings for all variables in the four Principal Components and % of explained variance.

280

The first principal component was negatively correlated with educational attainment (both for the
282 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
284 (Table 2). 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
286 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
288 and seventh component were mostly associated with salt and alcohol consumption, respectively, with
the rest of the variables showing lower loadings (Additional File 3).

290



292 **Fig. 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.

294 Density curves for Normal, Outlier and Extreme groups (bins=2) are superimposed.

296 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
298 the explanatory variables. Figure 2 (a) 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
300 head of the household). The higher density for the normal group was found in the region of higher

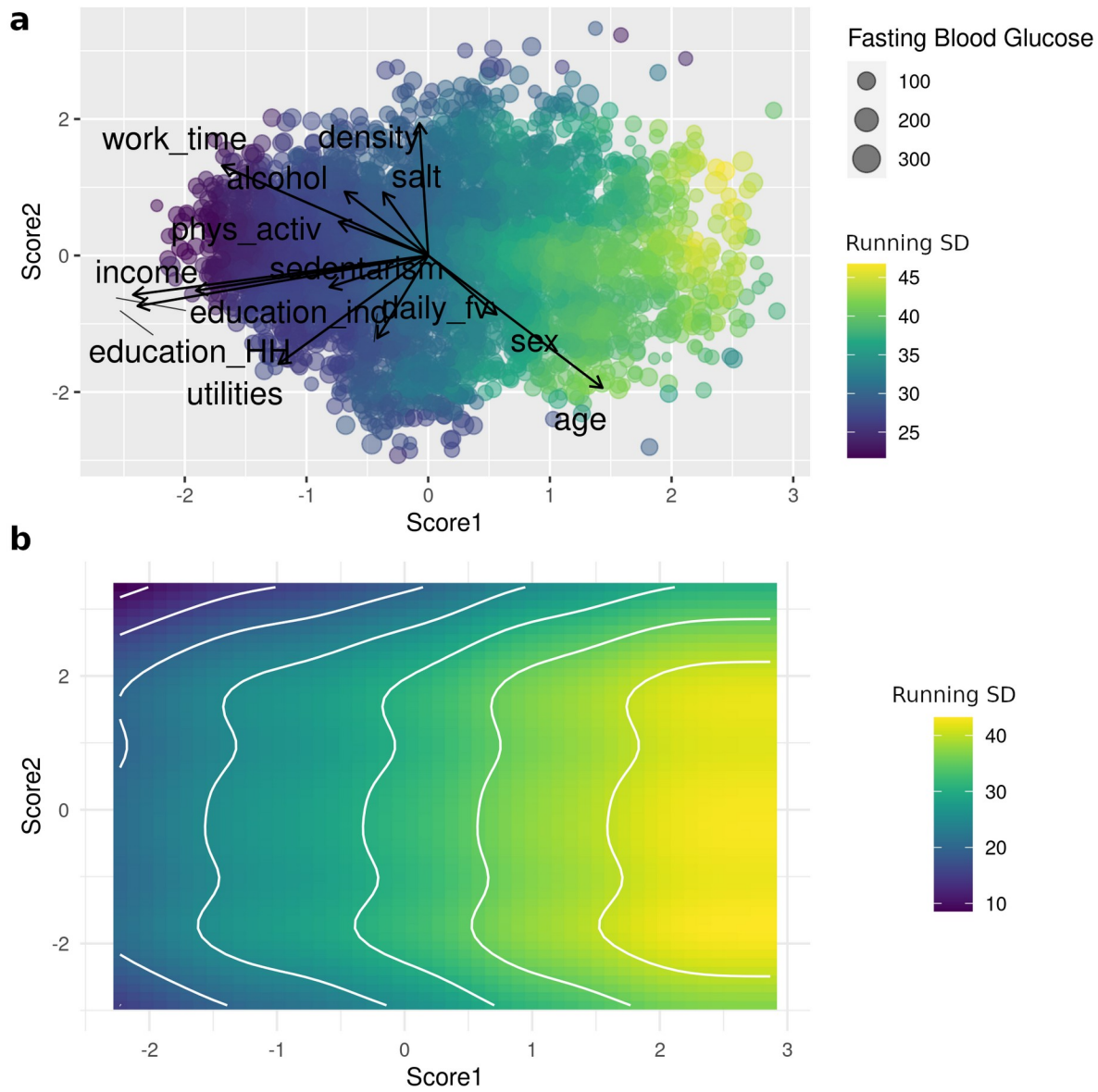
302 educational attainment, higher income and lower age. No clear pattern was found regarding lifestyle
variables, and Figure 2 (b) shows that the three groups mostly overlapped for the third and fourth
304 Principal Components. No clear patterns were found for the rest of the components (Additional File 4).
Similar results were obtained by using residuals of the linear regression between glycemia and age (not
306 shown).

308 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
310 performed a PERMANOVA, after confirming no significant departures from homogeneity of
multivariate dispersion in this multivariate space ($F_{2,4112} = 0.729$; $p = 0.47$). The PERMANOVA
312 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
314 the PCA ($F_{2,4112} = 20.95$; $p = 0.001$). Indeed, the three groups differed significantly in pairwise
comparisons (Table 3).

316

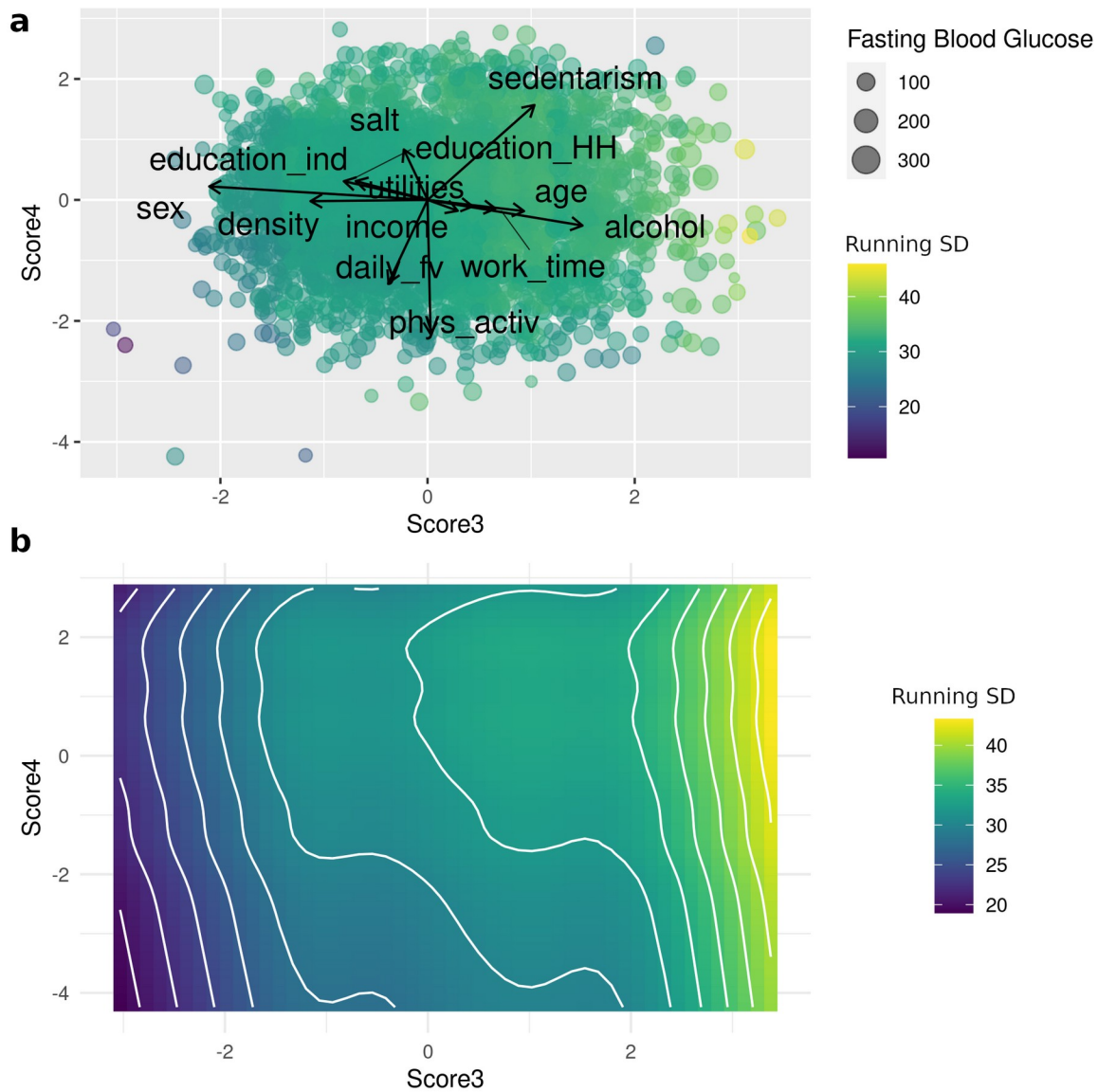
	F	Adjusted p-value
Normal - Extreme	7.22	0.003
Normal - Outlier	4.43	0.003
Extreme - Outlier	3.38	0.009

318 Table 3. Results of the pairwise comparisons for the PERMANOVA test



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

326



328

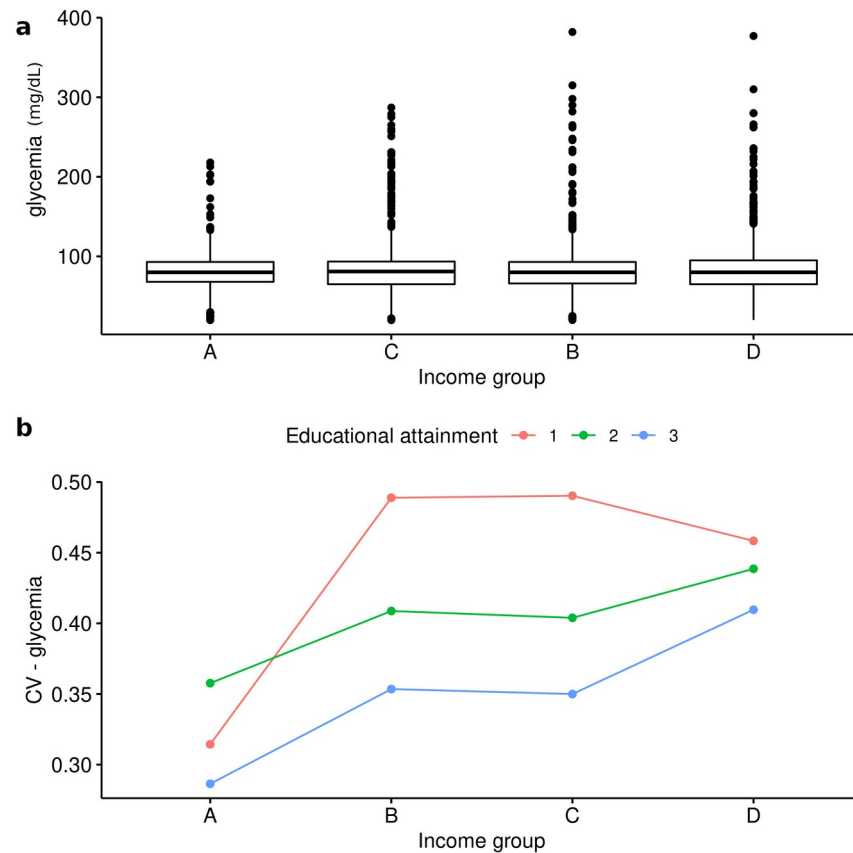
Fig. 4 Running standard deviation for fasting glycemia in the third and fourth components of the PCA

330 (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)

332 GAM smoothed curves for the running standard deviation.

334 The calculations of the running standard deviations showed a clear correlation between the first
principal component and heterogeneity for fasting blood glucose (Figure 3), with the regions
336 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
338 corresponding to the higher scores for the third component (higher alcohol consumption, older age,
male gender), but no clear tendency was apparent overall (Figure 4). No defined pattern was observed
340 either for the rest of the components (Additional File 5). These results did not change when correcting
by running average (not shown).

342



344 **Fig. 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
346 Variation for glycemia for the four income groups and the three levels of educational attainment.

As income and educational attainment appear as highly correlated relevant explanatory variables, we formed four income categories (see Materials and Methods) 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.



Fig. 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.

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

368 contributions to the statistic for each category can be found in Figure 6. For the high income and high
education group, individuals in the Normal category for the regression between glycemia and age were
370 more frequent than expected, while there was an excess of individuals at the Extreme and Outlier
categories in the lower income and lower education groups. Interestingly, this trend was reversed in the
372 low income and high education (C-3) group. However, we could not evaluate if this effect held for the
higher and lower income groups, where lower and higher educational attainment, respectively, were not
374 frequent.

376 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
378 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
380 that can be categorized as prediabetes, diabetes or severe diabetes (Figure 7). It may be noted that
hypoglycemia is also more prevalent in the lower income groups.

382

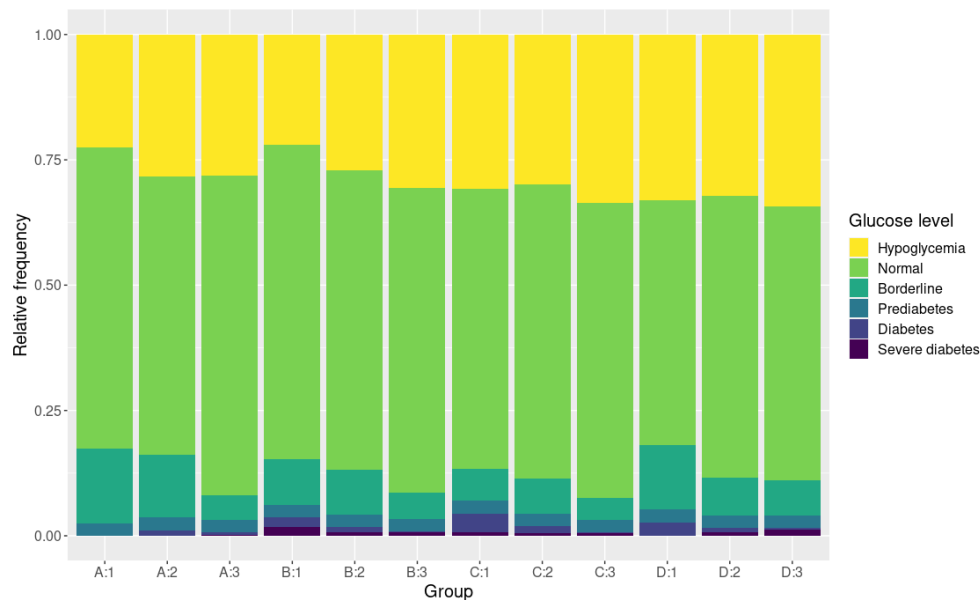


Fig. 7 Levels of fasting glycemia per income and educational attainment

384 Relative frequency of individuals presenting each glucose level at each income and educational
attainment combination

386

Discussion

388

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 best 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 [1, 3]. The data here analyzed show little to no evident relationship between lifestyle variables and variability for blood 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 [19].

398

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: predicting individual risk would be highly ineffective, as the augmentation of variance constitutes 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.

408

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 Figure 7. These high fasting blood sugar values lower quality of life, increase the risk of short and long-term complications, some even life-threatening [1], and worsen the prognosis for several other conditions [4, 5, 38].

414

416 There is a number of advantages of this kind of analysis. First, it overcomes some of the most frequent
critiques to traditional epidemiological research [39] such as that it does not propose mechanisms
418 behind increased prevalence, and therefore the real causes of ill health may become invisibilized. Also,
that it tends to disregard, minimize or simplify social and economic variables [40]. This is particularly
420 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)
422 factors, therefore making the individuals responsible for their ill health, and promoting constant self-
surveillance [41, 42]. In contrast, our approach includes a possible mechanism behind the increased
424 prevalence of complex disease in vulnerable (or vulnerated [43]) populations that can accommodate a
more holistic understanding of health and pathology processes, including the shifting social and
426 economic relationships in the explanation of phenotypic changes [40], and also would lead to concrete
strategies and public policy recommendations, without focusing on individual behaviors.

428 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
430 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
432 boundaries between health and disease can be seen as an advantage and an opportunity to leave open
the discussion on such boundaries.

434
in this sense, blood glucose determinations are an inestimable tool, as previous surveys only accounted
436 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
438 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.

440
However, our analysis has some disadvantages and limitations. First of all, biochemical determinations
442 were only carried out in interviewees from districts with a population higher than 150.000 [19].
Therefore, our conclusions can only apply to an urban population, as previously mentioned. Another
444 limitation on the survey concerns age, since all interviewees were adults (age \geq 18 years), although

446 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).

448 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).

452 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.

456

Another downside of this approach is the inability to distinguish between type 1, type 2, and gestational diabetes. However, as mentioned above, type 1 diabetes is usually detected in childhood or adolescence [44] and has a low prevalence, with an estimate for Argentinian children and adolescents of 8.6 per 100,000 people [45] and of 15 per 100,000 people worldwide [46], and is also more likely to depend on genetic factors [47]. 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 [6], and therefore it is expected to account for the majority of the cases recorded in the survey.

470 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 health outcomes can be pervasive in different groups [14, 48]. Here, in the case of educational attainment and its apparent interaction with family income, 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 we have seen in middle-income individuals with
476 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
478 increased social bonding, favoring autonomy and empowerment, and/or making it feasible to secure a
steady job and therefore more stable living conditions). Indeed, other researchers have already
480 proposed that access to higher education may counterbalance to some extent the heightened
susceptibility to ill health due to low income ([14] and references). Our results are consistent with both
482 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
484 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
486 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
488 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,
490 qualitative studies, to better understand the processes behind health disparities and how these are
experienced by individuals and communities in different contexts [49].

492
To our knowledge, this is the first work that addresses a possible relationship between these
494 socioeconomic variables and the decanalization hypothesis for fasting blood glucose. It is worth noting
that the conclusions reached regarding public policy are very similar to the ones proposed by Marmot
496 [14]: 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.)
498 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.
500 Himmelstein et al. [13], who found results consistent with the decanalization hypothesis for blood
pressure, suggest reducing the exposure to these stressors and protecting at-risk populations from
502 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.
504 Marmot [14] similarly argues that empowerment acts as a protecting factor and suggests policies to

improve income, education, housing stability, working conditions, etc. Our results support these
506 suggestions. Although these policies would imply an increase in spending not directly focused on
health care, such measures may have more meaningful impacts than current health policies for diseases
508 like type 2 diabetes, which constitute a growing economic burden both for individuals and health
systems [50].

510

We believe the decanalization hypothesis is a promising model for the study of the complex and
512 historically changing patterns of susceptibility to ill health, particularly to chronic multi-factorial
ailments related to homeostasis, which can incorporate biological (ecological, developmental,
514 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
516 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
518 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
520 [12, 39, 40, 49]. 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
522 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
524 overcoming disciplinary fragmentation. This issue has been raised for several subdisciplines of biology
[51] including those involved in human health [12, 40, 52, 53]. But moreover, our own notions and
526 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
528 revisions of our approaches and substantiate a more pluralistic take on these issues [49].

530 **Statements**

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

The authors declare they have no actual or potential competing financial interests. No specific funding
534 was destined to the preparation of this manuscript and the authors have no affiliations with or

involvement in any organization or entity with any financial interest or non-financial interest in the
536 subject matter or materials discussed in this manuscript.

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

540 **References**

1. World Health Organization. Diabetes. 2020. <https://www.who.int/health-topics/diabetes>. Accessed
542 December 2020.
2. Gibson G. It takes a genome: how a clash between our genes and modern life is making us sick. 1st
544 ed. FT Press; 2008.
3. World Health Organization. Global report on diabetes. 2016.
546 <https://www.who.int/publications/i/item/9789241565257>. Accessed May 2021.
4. Carrasco-Sánchez FJ, López-Carmona MD, Martínez-Marcos FJ, Pérez-Belmonte LM, Hidalgo-
548 Jiménez A, Buonaiuto V, Suárez Fernández C, Freire Castro SJ, Luordo D, Pesqueira Fontan PM,
Blázquez Encinar JC. Admission hyperglycaemia as a predictor of mortality in patients hospitalized
550 with COVID-19 regardless of diabetes status: data from the Spanish SEMI-COVID-19 Registry. *Ann
med.* 2021;53(1):103-16. <https://doi.org/10.1080/07853890.2020.1836566>
5. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology,
552 prognosis and practical considerations. *Diabetes Metab Syndr: Clinical Research & Reviews.*
2020;14(4):303-10. <https://doi.org/10.1016/j.dsx.2020.04.004>
6. World Health Organization. Prevention of Diabetes mellitus. 2014. Technical reports, Geneva
- 556 7. Waddington CH. Canalization of development and the inheritance of acquired characters. *Nature.*
1942;150(3811):563-5.
- 558 8. Flatt T. The evolutionary genetics of canalization. *Q Rev Biol.* 2005;80(3):287-316.
<https://www.journals.uchicago.edu/doi/abs/10.1086/432265>
- 560 9. Schmalhausen II. Factors of evolution: the theory of stabilizing selection. The Blakiston Co,
Philadelphia; 1949.
- 562 10. Gibson G. Decanalization and the origin of complex disease. *Nature Rev Genet.* 2009;10(2):134-
40. <https://doi.org/10.1038/nrg2502>

- 564 11. Lewontin R & Levins R. Schmalhausen's law. *Capitalism Nature Socialism*. 2000; 11(4), 103–
108. doi:10.1080/10455750009358943
- 566 12. Lewontin R & Levins R. *Biology under the influence: dialectical essays on ecology, agriculture,
and health*. Monthly Review Press; 2007.
- 568 13. Himmelstein DU, Levins R, Woolhandler S. Beyond our means: patterns of variability of
physiological traits. *Int J Health Serv*. 1990;20(1):115-24. <https://doi.org/10.2190%2FBKDL-N7DB->
570 [BDW8-DPYY](https://doi.org/10.2190%2FBKDL-N7DB-BDW8-DPYY)
14. Marmot M. *The health gap: the challenge of an unequal world*. Bloomsbury, London; 2015.
- 572 15. Cassel J. Psychosocial processes and “stress”: Theoretical formulation. *Int J Health Serv*.
1974;4(3):471-82. <https://doi.org/10.2190%2FWF7X-Y1L0-BFKH-9QU2>
- 574 16. Gibson G, Lacey KA. Canalization and robustness in human genetics and disease. *Ann Rev Genet*.
2020;54:189-211. <https://doi.org/10.1146/annurev-genet-022020-022327>
- 576 17. Reed LK, Williams S, Springston M, Brown J, Freeman K, DesRoches CE, Sokolowski MB,
Gibson G. Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila*
578 *melanogaster*. *Genetics*. 2010;185(3):1009-19. <https://doi.org/10.1534/genetics.109.113571>
18. Kodama K, Tojjar D, Yamada S, Toda K, Patel CJ, Butte AJ. Ethnic differences in the relationship
580 between insulin sensitivity and insulin response: a systematic review and meta-analysis. *Diabetes care*.
2013;36(6):1789-96. <https://doi.org/10.2337/dc12-1235>
- 582 19. Instituto Nacional de Estadística y Censos. 4° Encuesta Nacional de Factores de Riesgo. Resultados
definitivos. 2019.
584 https://www.indec.gob.ar/ftp/cuadros/publicaciones/enfr_2018_resultados_definitivos.pdf Accessed
May 2021.
- 586 20. Ministerio de Salud de la Nación Argentina. *Materiales para la población*. 2021.
<https://bancos.salud.gob.ar/bancos/materiales-para-poblacion>. Accessed May 2021.
- 588 21. Boyns DR, Crossley JN, Abrams ME, Jarrett RJ, Keen H. Oral glucose tolerance and related factors
in a normal population sample—I. Blood sugar, plasma insulin, glyceride, and cholesterol
590 measurements and the effects of age and sex. *Br Med J*. 1969;1(5644):595-8.
<https://doi.org/10.1136/bmj.1.5644.595>
- 592 22. Fletcher B, Gulanick M, Lamendola C. Risk factors for type 2 diabetes mellitus. *J Cardiovasc Nurs*.
2002;16(2):17-23.

- 594 23. Instituto Nacional de Estadística y Censos. Valorización mensual de la canasta básica alimentaria y
de la canasta básica total. Gran Buenos Aires. 2018.
596 https://www.indec.gob.ar/uploads/informesdeprensa/canasta_12_18.pdf. Accessed May 2021.
24. Allaire, J. RStudio: integrated development environment for R. Boston, MA, 2012; 770, 394.
598 <http://www.rstudio.com/>. Accessed May 2021.
25. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical
600 Computing, Vienna, Austria. 2020. <https://www.R-project.org/>
26. Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002;2(3):18-22.
- 602 27. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1-26.
<http://www.math.chalmers.se/Stat/Grundutb/GU/MSA220/S18/caret-JSS.pdf>. Accessed May 2021
- 604 28. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang CC, Lin CC, Meyer MD.
Package ‘e1071’. R J. 2019.
- 606 29. Breiman, L., Friedman, J.H., Olshen, R., & Stone, C. Classification and Regression Trees.
Wadsworth & Brooks. Monterey, CA; 2003.
- 608 30. Vapnik VN. An overview of statistical learning theory. IEEE T Neural Networ. 1999;10(5):988-99.
<https://doi.org/10.1109/72.788640>. Accessed May 2021.
- 610 31. Noble WS. What is a support vector machine?. Nat biotechnol. 2006;24(12):1565-7.
<https://doi.org/10.1038/nbt1206-1565>. Accessed May 2021.
- 612 32. Mair P, de Leeuw J. A general framework for multivariate analysis with optimal scaling: The R
package aspect. UCLA: Department of Statistics, UCLA. 2009.
614 <https://escholarship.org/uc/item/5q210583>. Accessed May 2021.
33. De Leeuw J, Mair P, Groenen PJF. Multivariate Analysis with Optimal Scaling. 2017.
616 <https://bookdown.org/jandeleeuw6/gif/ main.pdf>. Accessed May 2021.
34. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral ecol.
618 2001;26(1):32-46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
35. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’hara RB, Simpson GL, Solymos P,
620 Stevens MH, Wagner H, Oksanen MJ. Package ‘vegan’. Community ecology package, version.
2013;2(9):1-295. <https://CRAN.R-project.org/package=vegan>. Accessed May 2021.
- 622 36. Martinez Arbizu P. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package
version 0.4. GitHub. 2019. <https://github.com/pmartinezarbizu/pairwiseAdonis>. Accessed May 2021.

- 624 37. Signorell A. Package ‘DescTools’. Version 0.99. 32. 2020.
<https://CRAN.R-project.org/package=DescTools>. Accessed May 2021.
- 626 38. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related
complications. *Phys ther.* 2008;88(11):1254-64. <https://doi.org/10.2522/ptj.20080020>
- 628 39. Susser E. Eco-epidemiology: thinking outside the black box. *Epidemiology.* 2004;15(5):519-20.
doi: 10.1097/01.ede.0000135911.42282.b4
- 630 40. Breilh J. La determinación social de la salud como herramienta de transformación hacia una nueva
salud pública (salud colectiva). *Rev Fac Nac Salud Pub.* 2013;31:13-27.
- 632 41. Broom D, Whittaker A. Controlling diabetes, controlling diabetics: moral language in the
management of diabetes type 2. *Soc sci & med.*
634 2004;58(11):2371-82.<https://doi.org/10.1016/j.socscimed.2003.09.002>
42. Rail G, Jette S. Reflections on biopedagogies and/of public health: On bio-others, rescue missions,
636 and social justice. *Cultural Studies- Critical Methodologies.* 2015;15(5):327-336.
<https://doi.org/10.1177%2F1532708615611703>
- 638 43. Cunha T, Garrafa V. Vulnerability A Key Principle for Global Bioethics?. *Cambridge Q Healthc
Ethic.* 2016;25:197. <http://dx.doi.org/10.1017/S096318011500050X>
- 640 44. Chiang JL, Kirkman MS, Laffel LM, Peters AL. Type 1 diabetes through the life span: a position
statement of the American Diabetes Association. *Diabetes care.* 2014;37(7):2034-54.
642 <https://doi.org/10.2337/dc14-1140>
45. Patterson CC, Karuranga S, Salpea P, Saeedi P, Dahlquist G, Soltesz G, Ogle GD. Worldwide
644 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 Pr.* 2019;157:107842.
646 <https://doi.org/10.1016/j.diabres.2019.107842>
46. Mobasser M, Shirmohammadi M, Amiri T, Vahed N, Fard HH, Ghajzadeh M. Prevalence and
648 incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health promot
perspect.* 2020;10(2):98. <https://dx.doi.org/10.34172%2Fhpp.2020.18>
- 650 47. Borchers AT, Uibo R, Gershwin ME. The geoepidemiology of type 1 diabetes. *Autoimmunity rev.*
2010;9(5):A355-65. <https://doi.org/10.1016/j.autrev.2009.12.003>
- 652 48. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature.
Circulation. 1993;88(4):1973-98. <https://doi.org/10.1161/01.CIR.88.4.1973>

- 654 49. Almeida Filho, N. Epidemiología sin Números. Washington DC: Organización Panamericana de la
Salud; 1992.
- 656 50. Bommer C, Sagalova V, Heeseemann E, Manne-Goehler J, Atun R, Bärnighausen T, Davies J,
Vollmer S. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes care.
658 2018;41(5):963-70. <https://doi.org/10.2337/dc17-1962>
51. Alberti M, Marzluff JM, Shulenberger E, Bradley G, Ryan C, Zumbrunnen C. Integrating humans
660 into ecology: opportunities and challenges for studying urban ecosystems. BioScience.
2003;53(12):1169-79. [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)
- 662 52. Schölmerich VL, Kawachi I. Translating the socio-ecological perspective into multilevel
interventions: Gaps between theory and practice. Health Educ Behav. 2016;43(1):17-20.
664 <https://doi.org/10.1177%2F1090198115605309>
53. Levins R, Dunn ML. Whose scientific method? Scientific methods for a complex world. New
666 Solut. 2003;13(3):261-74. doi:10.2190/q4tn-q9u2-er56-3t1r

668

670 **Additional file captions**

Additional File 1. Data from the National Health Survey for all individuals retained.

672 Additional File 2. Summary of the regressions performed to model blood glycemia data. Method,
proportion of variance explained and the detailed models (in R language) are provided

674 Additional File 3. Loadings for all the variables in the 8 principal components generated

Additional File 4. Curves for Normal, Outlier and Extreme glycemia residuals. Biplots for the
676 components 5, 6, 7 and 8 of the PCA. Density curves for Normal, Outlier and Extreme groups (bins=2)
are superimposed.

678 Additional File 5. Running standard deviation for fasting glycemia. Biplots for the components 5, 6, 7
and 8 of the PCA; point size correlates with fasting blood glucose levels (mg/dL) and color scale
680 indicates levels of running standard deviation (mg/dL). GAM smoothed curves for the running standard
deviation are also provided.

682