Acid Phosphatase, some Genetic Polymorphism and obesity risk factors in adult women

E. Carolino
EST&SL – IPL.
Laboratório de genética – FML-UL.
lizcarol@iol.pt

M. R. Ramos
Universidade Aberta.
CMAF-UL.
ramos@cii.fc.ul.pt

A. P. Silva
Laboratório de genética – FML-UL.

R. Carvalho
Laboratório de genética – FML-UL.

M. Bicho
Laboratório de genética – FML-UL.

Abstract: Recent works point out to a relation between some genetic factors and the predisposition for obesity. We believe, therefore, to be relevant to conduct this kind of study in the Portuguese population. In the present work the following genetic factors are considered: Haptoglobin phenotype, the Acid Phosphatase phenotype and two blood group systems, the MN System and the Lewis System. In addition, it was also considered one demographic factor, age, and one enzymatic activity, the Acid Phosphatase Activity.
Haptoglobin (Hp) is a hemoglobin-binding protein of the immune system expressed by a genetic polymorphism with three major phenotypes. This protein is associated in some works with susceptibility for common pathological situations, such as some disorders related with obesity.
The Acid phosphatase, more precisely the Acid phosphatase locus 1 (ACP1), is a highly polymorphic enzyme that has an important role in flavoenzyme activity and in the control of insulin receptor activity. High ACP1 activity was positively associated with high glycemic levels and with high body mass index (BMI) values.
The MN blood system is a blood group system with three phenotypes each one showing different associations with some diseases, including some related with obesity.
Finally, the Lewis System was focused on a single locus with two antigens, Le a and Le b. Confirming this characteristic as a genetic marker of obesity may contribute to the explanation of individual differences in the prevalence of obesity.
The group under study involves 85 Portuguese adult women with complete data for all variables, taken from a data base with 714 subjects from the Genetic Laboratory

— Centre of Endocrinology and Metabolism of University of Lisbon.
The aim of the study is to explore and examine the relationship between the weight categories and the explanatory variables, with emphasis on risk for obesity. Therefore, an ordinal regression model was tried, considering as the regressor variables the Haptoglobin phenotype, Acid phosphatase (ACP1) phenotype, MN blood group system, Lewis system, the enzymatic activity of ACP1, age and some association effects between these factors.
Some significant main effects were found at a 5% significance level: the phenotype Le(a-b+) of Lewis System (*p-value*=0,021) and age (*p-value*=0,002). The phenotype Le(a-b+) of Lewis System is associated with a decreased risk for obesity (odds ratio 0,139; $\mathrm{CI}_{95\%}(0{,}016;\ 0{,}754)$); age (as expected) is associated with an increased risk for obesity (odds ratio 1,11; $\mathrm{CI}_{95\%}(1{,}038;\ 1{,}190)$).

Keywords: Obesity; Lewis blood group; MN blood group; Ordinal Regression; Probit; odds ratio

1 Introduction

One of the most current issues in epidemiological studies is the search for factors associated with public health problems which have increased alarmingly in developed countries, in which obesity is included. The recent discoveries in human genetics have fostered the development of intervention of individualized strategies (prevention and/or treatment).
In this context the present work aims to identify individual characteristics that may be classified as risk factors. It is known from previous studies that there are some genetic factors related with the predisposition for obesity. In the present, some potential risk factors for obesity are studied, namely, four genetic factors such as two different blood group systems, the Lewis System and MN System, the Haptoglobin phenotypes, the Acid Phosphatase phenotypes, one particular enzymatic activity, the Acid Phosphatase activity, and one demographic factor, the age.
Haptoglobin (Hp) is a protein of the immune system which exists in two allelic forms in the human population, so called Hp1 and Hp2. Three phenotypes of Hp therefore are found in humans: Hp1-1, Hp2-1, and Hp2-2. This protein is associated with the susceptibility for common pathological situations, such as obesity.
The Acid phosphatase, or ACP1, is an intracellular enzyme that posses phosphotyrosine phosphatase and flavin-mononucleotide (FMN) phosphotransferase protein activity. This gene is genetically polymorphic, and three common alleles segregating at the corresponding locus give rise to six phenotypes: AA, AB, AC, BB, BC and CC. Previous works by the authors have shown some evidence of an association between obesity and the activity levels of the Acid phosphatase.
MN is a blood group system in which two alleles identified as M and N determine the presence of corresponding antigens on the red cells and give rise to three phenotypes MM, MN and NN. Several disease associations involving these phenotypes have surfaced recently, including associations with cardiovas-

cular diseases, environmental induced hyperlipidemia and hypercolestorelemia, all somehow related with obesity.
The Lewis System was focused on a single locus with two antigens, **Le *a*** and **Le *b***. Because having both antigens is extremely rare, there are only three phenotypes in the study: Le(a-b-); Le(a+b-); and Le(a-b+). The small number of studies on disease association, particularly in obesity, justifies the choice of this blood group.
The Body Mass Index (BMI) provides an indicator of body fatness that according to the World Health Organization is classified in four categories of weight (underweight, normal weight, overweight and obese).

2 Material and Methods

The study sample is composed of 85 Portuguese adult women, aged between 18 and 98 years old, with complete data for all the variables mentioned previously. Data were taken from a data base with 714 subjects of the Genetic Laboratory — Centre of Endocrinology and Metabolism of University of Lisbon.
The weight status is often measured on an ordinal scale. Considering four weight categories an Ordinal Regression Model was design in order to investigate the relationship among the regressors and the risk for obesity. The results may be interpreted in terms of odds ratios.
Generalized linear models for ordinal data include the ordinal regression model. The model is based on the assumption that there is a latent continuous outcome variable, in this case BMI, and that the observed ordinal outcome arises from discretizing the underlying continuum into i-ordered groups (BMI categories) on which one assigns the response variable Y. The thresholds estimate these cut-off values.
As regressor variables in the final model it were chosen the Lewis blood group system (X_1), the MN blood group system (X_2), Haptoglobin genotype (X_3), Acid phosphatase phenotype (X_4), age (X_5) and the Acid phosphatase activity levels (X_6). The model also includes the interaction effects X_4X_6.
The basic form of the generalized linear general model is

$$link(\gamma_i) = \frac{\alpha_i - [\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p]}{e^{(a_1 s_1 + a_2 s_2 + ... + a_k s_k)}}, i = 0, 1, 2, 3 \tag{1}$$

The numerator models the location; the denominator, the scale. Note that, if all $a_j = 0$ $(j = 1, ..., k)$ the model reduces to a location model.
In the matrix form

$$\gamma_i = Link\left(\frac{\alpha_i - \boldsymbol{\beta}\mathbf{X}}{e^{\mathbf{as}}}\right), i = 0, 1, 2, 3 \tag{2}$$

where:

- γ_i represents the probability that the event i occurs. In ordinal regression, the probability of an event is redefined in terms of cumulative probabilities,

- α_i represents a separate intercept or threshold for each cumulative probability — unknown parameter to be estimated by means of the maximum likelihood method;

- $\boldsymbol{\beta}$ is the vector of regression coefficients — unknown parameters to be estimated by means of the maximum likelihood method;

- $\mathbf{X}$ is the design matrix of the regressor variables and p the number of predictors;

- $\mathbf{a}$ is the vector of coefficients for the scale component;

- $\mathbf{s}$ is the vector of the predictor variables in the scale component (chosen from the same set of variables in $\mathbf{X}$).

The scale component accounts for differences in variability for different values of the predictor variables.
The Ordinal Regression model with the "*Probit*" link function for ordered categorical dependent variable seems to be the most appropriate choise for this set of data, since the latent variable, BMI, is approximately normally distibuted[1]. The link function is then

$$\gamma_i = P(Y \leq i|\mathbf{X}) = \Phi\left(\frac{\alpha_i - \boldsymbol{\beta}\mathbf{X}}{e^{\mathbf{as}}}\right), i = 0, 1, 2, 3 \tag{3}$$

being Φ the standardized normal distribution. The model can rewritten as

$$\Phi^{-1}(P(Y \leq i|\mathbf{X})) = \frac{\alpha_i - \boldsymbol{\beta}\mathbf{X}}{e^{\mathbf{as}}}, i = 0, 1, 2, 3 \tag{4}$$

where Φ^{-1} is the inverse of the cumulative normal distribution.
Some regresor variables are categorical with more than two categories, so it was necessary to create dummy variables for each one. Conceptually

$$X_{ik} = \begin{cases} 1 & \text{, in the presence of the category } k \\ 0 & \text{, otherwise} \end{cases}, k = 1, ..., nb \text{ categories} - 1 \tag{5}$$

The reference category is the last one and i represents categorical variable i.

According to this formulation, we have the linear component of the general

---

[1] Several studies report the approximate normality of BMI. However, more recent studies show some bias to the right. For the Portuguese population and for women in particular were not found studies available. So, the authors assume normality, since the exploratory analysis does not show a significant bias.

model:

$$\boldsymbol{\beta}\mathbf{X} = \sum_{i=1}^{2}\beta_{1i}X_{1i} + \sum_{j=1}^{2}\beta_{2j}X_{2j} + \sum_{k=1}^{2}\beta_{3k}X_{3k} + \sum_{l=1}^{5}\beta_{4l}X_{4l} + \beta_5 X_5 + \beta_6 X_6 + \sum_{l=1}^{5}\beta_{7l}X_{4l}X_6 \qquad (6)$$

When a scale parameter is present, the null hypothesis states that both the location and scale parameters are 0. A scale parameter of 0 means that the variances are equal. Thus, to perform an overall test of null hypothesis that all coefficients of the variables in the model are 0, i.e. $H_0 : \boldsymbol{\beta} = \mathbf{0}$, and $\mathbf{a} = \mathbf{0}$, we use the change -2log-likelihood when the variables are added to a model that contains only the intercept. The change in likelihood function is distributed as a Chi-square, with $p+k-1$ degrees of freedom, where p is the number of coefficients of the location and k is the number of coefficients in the scale component of the model.
The parameters α_i (coefficients of the thresholds), β_j (coefficients of the variables in location component) and a_l (coefficients in the scale component) were estimated by maximizing the likelihood induced by the model in (3), and the inference effected using the Wald test, in which the statistic $\left[\frac{\hat{\beta}-\beta_0}{SE(\hat{\beta})}\right]^2$ is qui-squared distributed with a number of degrees of freedom equal to the number of categories minus 1.
To compare the full model, including p predictors, with the corresponding reduced model, excluding the q predictors related to non-significant coefficients, we used the Likelihood ratio test. For this test we have:

$$G = -2[\log likelihood(\text{reduced model}) - \log likelihood(\text{full model})] \qquad (7)$$

Under the null hypothesis, that the coefficients of the excluded variables are zero, the statistic G follows a Chi-square distribution with pq degrees of freedom.

Whole the computation work was performed with the statistical package SPSS 17.0.

3 Results

After a careful exploratory data analysis, it was found that the frequency distribution of weight categories differ between phenotypes of the MN blood group system and with Haptoglobin phenotypes, as illustrated in the figures 1 and 2. We are interested, then, in the study of an ordinal regression model with scale parameters, considering as regressors those defined in (6) and as scale parameters the MN blood group system and Haptoglobin phenotypes.

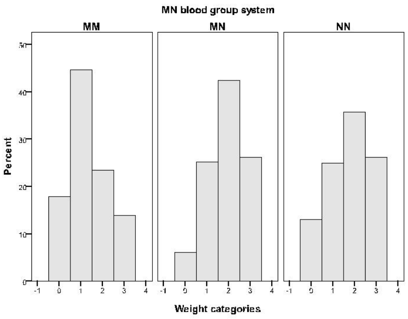

Figure 1: Frequency distribution of weight categories by blood group MN

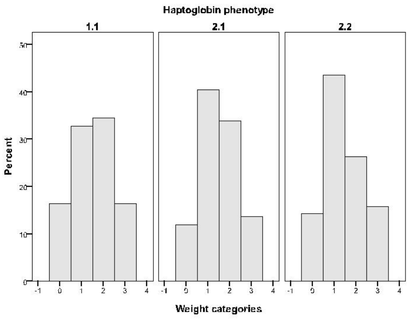

Figure 2: Frequency distribution of weight categories by Haptoglobin phenotype

Considering the model defined in (3) and (6), for which $-2 loglikelihood =$ 155,840 and *p-value*=0,000, the coefficients of the acid phosphatase were not significant. It was then fitted another model removing this variable, for which $-2 loglikelihood = 161{,}496$ and *p-value*=0,000. In both, cases, the significance test of the model, based on statistic G (equation 7), leads to the rejection of the null hypothesis that all coefficients of the variables in the model are zero. Comparing the full model with the reduced one, the test statistic assumes the value G =5,656, concluding that there is no significant improvement in the model quality. Despite this result, for reasons of parsimony and to overcome the problem of over-parameterization, a simpler model was chosen, putting aside the phenotype of the Acid phosphatase.

The results of the fit of the model are now shown. Both goodness-of-fit statistics, Pearson and Deviance, show evidence of an appropriate model (*p-value*>0,05). Additionally, the goodness-of-fit was assessed by the cross-tabulation between the observed and the predicted response categories. The analysis is based on the comparison of the percentage of corrected classified subjects by the model, with the percentage of corrected classified subjects by chance.

Taking the global classification, in table 1, the model classifies correctly 55,3% of the subjects. Comparing this with the percentage of correctly classified subjects by chance, only 27,8%, it is concluded that the model improves the classification. Moreover, it is possible to verify that the model predicts correctly 41,7% of the Underweight category, 80% of the Normal Weight category, 46,2% of the Overweight category and 35,3% of the Obese.

Cohen's Kappa statistics corroborates the previous results, ie, kappa=0,4 (*p-value*=0,000) with a 95% confidence interval (0,26; 0,54), indicating that there is a reasonable agreement in the classification of individuals.

The pseudo R-square value indicates an association between the dependent and predictor variables. The McFadden statistics assumes the value 0,29, Cox and Snell statistics equal to 0,51 and Nagelkerke statistics equal to 0,571.

Weight categories * Predicted Response Category Cross-tabulation

			Predicted Responde Category				Total
			Under-weight	Normal weight	Over-weight	Obese	
Weight categories	Under-weight	Count	5	6	0	1	12
		% within Weight categories	41,7%	50,0%	0%	8,3%	100,0%
	Normal weight	Count	2	24	1	3	30
		% within Weight categories	6,7%	80,0%	3,3%	10,0%	100,0%
	Over-weight	Count	1	7	12	6	26
		% within Weight categories	3,8%	26,9%	46,2%	23,1%	100,0%
	Obese	Count	2	1	8	6	17
		% within Weight categories	11,8%	5,9%	47,1%	35,3%	100,0%
Total		Count	10	38	21	16	85
		%within Weight categories	11,8%	44,7%	24,7%	18,8%	100,0%

Table 1: Results of cross-tabulation between observed weight categories and predicted response weight categories

Table 2 shows the results from the maximum likelihood algorithm. The significant coefficients in the location component are the phenotype Le(a-b+) of Lewis System (*p-value*=0,021) and age(*p-value*=0,002). The significant scale coefficients are phenotype MN (*p-value*=0,042) of MN Blood System and the phenotype Hp1-1 of Haptoglobin (*p-value*=0,023).

For every increase of ten years in age, the risk of belonging to a higher weight category increases 11,1 times ($\widehat{OR}$ = 1,11, CI$_{95\%}$=(1,04; 1,19)). Associated with a decrease in risk for obesity is the phenotype Le(a-b+) of Lewis System ($\widehat{OR}$=0,14, CI$_{95\%}$=(0,02; 0,75)) in respect to the reference category phenotype a-b-.

Parameter Estimates

		Estimate	Std Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	Underweight	,096	,731	,017	1	,896	-1,338	1,530
	Normal weight	2,337	1,035	5,099	1	,024	,308	4,365
	Overweight	3,782	1,365	7,678	1	,006	1,107	6,457
Location	a+b-	-,204	,338	,365	1	,546	-,867	,459
	a-b+	**-1,167**	**,505**	**5,339**	**1**	**,021**	**-2,156**	**-,177**
	a-b-	0[a]	.	.	0	.	.	.
	MM	,000	,409	,000	1	1,000	-,801	,801
	MN	,159	,365	,191	1	,662	-,556	,874
	NN	0[a]	.	.	0	.	.	.
	Hp1-1	,139	,623	,050	1	,823	-1,082	1,361
	Hp2-1	,357	,301	1,409	1	,235	-,233	,947
	Hp2-2	0[a]	.	.	0	.	.	.
	age	**,066**	**,022**	**9,273**	**1**	**,002**	**,024**	**,109**
	Acid Phosphatase activity	-,002	,004	,344	1	,557	-,010	,005
	AA*Acid Phosphatase activity	,001	,004	,046	1	,829	-,007	,008
	BB*Acid Phosphatase activity	,000	,003	,003	1	,960	-,006	,007
	CC*Acid Phosphatase activity	,003	,004	,620	1	,431	-,004	,010
	AB*Acid Phosphatase activity	-,001	,003	,105	1	,746	-,008	,006
	AC*Acid Phosphatase activity	-,001	,004	,027	1	,871	-,007	,006
	BC*Acid Phosphatase activity	0[a]	.	.	0	.	.	.
Scale	MM	,361	,282	1,646	1	,199	-,191	,913
	MN	**-,728**	**,359**	**4,119**	**1**	**,042***	**-1,431**	**-0,025**
	NN	0[a]	.	.	0	.	.	.
	Hp1-1	**,806**	**,355**	**5,147**	**1**	**,023***	**,110**	**1,502**
	Hp2-1	-,050	,265	,036	1	,849	-,570	,469
	Hp2-2	0[a]	.	.	0	.	.	.

Link function: Probit
* significant at 5% significance level.
[a] estimates are not evaluated because it is redundant.

Table 2: Parameter estimates, odds ratios and confidence interval estimates

There are only two significant cut-off points (or thresholds), the threshold between the normal weight category and the overweight category and the threshold between the overweight category and the obese. This means that with this model and these explanatory variables, the sample data does not show a clear separation between the two lower categories, Underweight and Normal weight.

4 Concluding remarks

The use of statistical models which respects the ordinal nature of the response variable is preferable, because they contain more information than the alternatives that do not consider it. With this in mind, we sought a model that helps to identify risk factors for obesity. The search was guided not only by the significance of the regressor variables but also to obtain a parsimonious model. The final model has the regressors Lewis System, MN Blood System, Haptoglobin phenotype, age, Acid phosphatase activity and the interaction between the Acid phosphatase activity and the phenotypes of this enzyme. The MN Blood System, and the Haptoglobin phenotype are in the scale component.
The phenotype Le(a-b+) is associated with a decreased risk for obesity compared with the reference category Le(a-b-), being this decrease about 86%. In what concerns to the age there is an increase of 11% for obesity for each period of ten years.
Based on the results, the authors think that some of the chosen as potential risk factors can be very important to predict risk for overweight and obesity. Nevertheless, it seems important to carry out further studies and collect more data in order to confirm some results and to explore the potential relationship between overweight and obesity with other explanatory variables.

Bibliography

[1] Agresti, A. (1990). *Categorical Data Analysis.* 2nd ed., John Wiley & Sons.

[2] Armstrong, B. G. and Sloan, M. (1989). Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, 129(1), 191-204.

[3] Cole T. J., Freeman J. V. and Preece M. A. (1990). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood.*Statistics in Medicine*, 17(4), 407-429.

[4] David G. and Klein Michel (2002). *Logistic Regression.* 2nd ed. Springer.

[5] Giblett, E. R. (1969). *Genetic Markers in Human Blood.* Blackwell Scientific.

[6] Guerra, A *et al.* (2000). *Revista Portuguesa de Cardiologia*, 19(6), 679-91.

[7] Guerra, A *et al.* (2001). *Int. J. Obes. & Related Metabolic Disorders*, 25(3), S12.

[8] Gutteridge, J. M. (1987). The antioxidant activity of haptoglobin towards haemoglobin-stimulated lipid peroxidation. *Biochim. Biophys. Acta*, Vol. 917, pp. 219-223.

[9] Hosmer D. W. and Lemeshow, S. (2000). *Applied Logistic Regression.* 2nd ed. John Wiley & Sons.

[10] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd ed., Chapman and Hall.

[11] Mourant AE and Kopec AC. (1978). *Blood group and disease.* Oxford: Oxford University Press, London.

[12] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, 42(2), 109-142.

[13] SPSS 17.0. *http://www.spss.com/spss/whats_ new_ base.htm.*