Biometrical Letters
Vol. 46 (2009), No. 1, 43-54

Haptoglobin, acid phosphatase and demographic factors: obesity risk

M.R. Ramos[1,2], E. Carolino[3,4], T. Oliveira[1,5], A.P. Silva[4], R. Carvalho[4], M. Bicho[4]

[1]U. Aberta, Dept. Sciences and Technology, R. Fernão Lopes, 9, 2D, 1000-132 Lisbon, Portugal, toliveir@univ-ab.pt
[2]CMAF, Centre of Mathematics and Fundamental Applications, University of Lisbon, Portugal, ramos@cii.fc.ul.pt
[3]ESTeSL, Dept. Natural and Exact Sciences, Av. D. João II, Lote 4.69.01, 1990 - 096 Lisbon, Portugal, lizcarolino@gmail.com
[4] Genetics Laboratory – FML-UL, Lisbon, Portugal
[5] CEAUL, Centre of Statistics and Applications, University of Lisbon, Portugal

SUMMARY

The aim of this work is to study the risk of obesity posed by two genetic factors: haptoglobin phenotype and acid phosphatase phenotype, one enzymatic activity: acid phosphatase activity (ACP1), age and gender.
Haptoglobin (Hp) is a protein of the immune system, and three phenotypes of Hp are found in humans: Hp1-1, Hp2-1, and Hp2-2. This protein is associated with a susceptibility to common pathological conditions, such as obesity. ACP1 is an intracellular enzyme The phenotypes of ACP1 (AA, AB, AC, BB, BC, CC) are also considered.
We took a sample of 127 subjects with complete data from 714 registers. Since we intend to identify risk factors for obesity, an ordinal regression model is adjusted, using the Body Mass Index, BMI, to define weight categories. Haptoglobin phenotype, enzymatic activity of ACP1, acid phosphatase phenotype, age and gender are considered as regressor variables.
We found three factors associated with an increased risk of obesity: phenotype Hp2-1 of haptoglobin (estimated odds ratio $\widehat{OR}$ 11.54), phenotype AA of acid phosphatase ($\widehat{OR}$ 33.788) and age ($\widehat{OR}$ 1.39). The interaction between phenotype Hp2-1 and phenotype AC is associated with a decreased risk of obesity ($\widehat{OR}$ 0.032); The interaction between phenotype AA and ACP1 activity is associated with a decreased risk of obesity ($\widehat{OR}$ 0.954).

Key Words: Obesity, acid phosphatase, haptoglobin, ordinal regression, odds ratio

1. Introduction

Obesity is currently a serious problem of public health, so it is relevant to design research in order to identify individual characteristics that may be risk factors, and to establish some prevention strategies. It is known from previous studies that there are some genetic factors related associated with a predisposition to obesity.

The Body Mass Index (BMI) provides an indicator of body fatness, and according to the World Health Organization it enables classification in four categories of weight (underweight, normal weight, excess weight and obese).

In this work we study some potential risk factors, namely: two genetic factors, haptoglobin phenotypes and acid phosphatase phenotypes; one particular enzymatic activity: acid phosphatase activity; and demographic factors, such as age (between 18 and 98 years) and gender.

Haptoglobin (Hp) is a protein of the immune system. Hp exists in two allelic forms in the human population, called Hp1 and Hp2, the latter having arisen due to partial duplication of the Hp1 gene. Three phenotypes of Hp are therefore found in humans: Hp1-1, Hp2-1 and Hp2-2. Hp of different genotypes has been shown to bind haemoglobin with different affinities, with Hp2-2 being the weakest binder. This protein is associated with susceptibility to common pathological conditions, such as obesity.

The acid phosphatase ACP1 is an intracellular enzyme that has phosphotyrosine phosphatase and flavin-mononucleotide (FMN) phosphor-transferase protein activity. The phenotypes of acid phosphatase are AA, AB, AC, BB, BC and CC.

Considering weight categories (the response variable) as categories of an ordinal categorical variable, since there is a natural ordering of its possible values (underweight, normal weight, overweight and obese), an Ordinal Regression Model was developed in order to predict the risk of obesity. As regressor variables, we consider the enzymatic activity of acid phosphatase (ACP1 activity), acid phosphatase phenotype, haptoglobin phenotype, age and gender.

The adjusted model analysis leads to the conclusion that some of the selected risk factors may be important in predicting obesity risk. At a significance level of 5%, the most significant results associated with an increased risk of obesity were haptoglobin phenotype Hp2-1 (estimated odds

ratio $\overline{OR}$ 11.54; 95% confidence interval $CI_{95\%}$, (2.10; 21.74)), acid phosphatase phenotype AA ($\overline{OR}$ 33.79; $CI_{95\%}$ (7.73; 60.10)), and age ($\overline{OR}$ 1.39; $CI_{95\%}$ (1.15; 1.67)). The interaction between phenotype Hp2-1 and phenotype AC of acid phosphatase is associated with a decreased risk of obesity ($\overline{OR}$ 0.032; $CI_{95\%}$ (0.02; 0.11)) and the interaction between phenotype AA of acid phosphatase and the enzyme activity is also associated with a decreased risk of obesity ($\overline{OR}$ 0.95; $CI_{95\%}$ (0.92; 0.99)).

2. Material and methods

In this work we took a subset of 127 subjects from a database of 714 subjects from the Genetics Laboratory of the Centre of Endocrinology and Metabolism at the University of Lisbon, with complete data for all variables chosen for this study. The subjects are men and women aged between 18 and 98 years.

The four categories of weight were coded as follows: 0 for underweight; 1 for normal weight, 2 for excess weight, 3 for obese.

The three phenotypes of haptoglobin (Hp) were coded as follows: 0 for Hp1-1, 1 for Hp2-1, 2 for Hp2-2.

The acid phosphatase phenotypes were coded 0 for AA, 1 for AB, 2 for AC, 3 for BB, 4 for BC and 5 for CC.

The ordinal regression model is one of many models subsumed under the rubric of generalized linear models for ordinal data. The model is based on the assumption that there is a latent continuous outcome variable (BMI in the present study) and that the observed ordinal outcome arises from discretizing the underlying continuum into *i*-ordered groups (BMI categories, Y). The thresholds estimate these cut-off values. Thus we assume an ordinal regression model to predict the risk of obesity, considering as regressor variables haptoglobin phenotype (X_1), acid phosphatase phenotype (X_2), enzymatic activity of ACP1 (X_3), age (X_4) and gender (X_5). The model included the interaction effects X_1*X_2, X_2*X_3, X_3*X_4, X_4*X_5.

The Cauchit (inverse Cauchy) link function seems to be the most appropriate choice, since the outcome includes many extreme values.

This link function is defined by

$$\tan(\pi(\gamma_i - 0.5)) = \alpha_i - \boldsymbol{\beta}\mathbf{X} \quad \text{for } i=0, 1, 2, 3, \tag{1}$$

where γ_i is the cumulative probability for the *i*th category, α_i represents a separate intercept or threshold for each cumulative probability; **β** is the row vector of regression coefficients and **X** is the design matrix of the regressor variables. The threshold (α_i) and the regression vector of coefficients (**β**) are unknown parameters to be estimated using the maximum likelihood method. Note that, in ordinal regression, the probability of an event is redefined in terms of cumulative probabilities.

Thus the model equation is

$$\gamma_i = P(Y \le i|\mathbf{X}) = \frac{\tan^{-1}(\alpha_i - \boldsymbol{\beta}\mathbf{X})}{\pi} + 0.5 \quad , i=0, 1, 2, 3. \tag{2}$$

Some regressor variables are categorical with more then two categories, so it was necessary to create dummy variables for each one of these variables. For haptoglobin phenotype, two dummy variables were considered, namely

$$X_{1k} = \begin{cases} 1 \text{ , in the presence of phenotype } k \\ 0 \text{ , otherwise} \end{cases} \quad , k=0, 1, \tag{3}$$

being the reference category Hp2-2.

For the acid phosphatase phenotype five dummy variables were defined, as follows:

$$X_{2j} = \begin{cases} 1 \text{ , in the presence of phenotype } j \\ 0 \text{ , otherwise} \end{cases} \quad , j=0, 1, 2, 3, 4, \tag{4}$$

being the reference category for the phenotype BC.

According to this formulation, we have

$$\boldsymbol{\beta}\mathbf{X} = \sum_{i=0}^{1} \beta_{1i} X_{1i} + \sum_{j=0}^{4} \beta_{2j} X_{2j} + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \sum_{i=0}^{1}\sum_{j=0}^{4} \beta_{ij} X_{1i} X_{2j} + \sum_{j=0}^{4} \beta_{2j} X_{2j} X_3 + \beta_6 X_3 X_4 + \beta_7 X_4 X_5 \tag{5}$$

To perform an overall test of the null hypothesis that all coefficients of the variables in the model are null, i.e. $H_0 : \boldsymbol{\beta} = \mathbf{0}$, we used the change in -2logLikelihood 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–1 degrees of freedom, where p is the number of coefficients in the model.

To test the significance of individual coefficients, the Wald statistic was used, which under the null hypothesis is normally distributed.

To compare the full model, including p variables, with the corresponding reduced model, excluding the q variables related to non-significant coefficients, we used the Likelihood ratio test. For this test we have:

$$G = -2\left[loglikelihood\left(\text{reduced model}\right) - loglikelihood\left(\text{full model}\right)\right]$$

Under the null hypothesis that the coefficients of the excluded variables are zero, the statistic G follows a chi-square distribution with p–q degrees of freedom.

The statistical package SPSS 16.0 was used for the statistical data analysis.

3. Results

After the adjustment of the postulated model as formalized in equations (2) and (5), the goodness-of-fit statistic indicates that the model, henceforth called the full model, does not fit well (the chi-square statistic with 351 d.f. is 421.699, and the corresponding *p-value*=0.006<0.05).

Given this results for the full model, we tried a reduced model, removing the regressor “gender” from the equation. The results of the goodness-of-fit tests lead to a good fit (the chi-square statistic with 352 d.f. is 386.295 and the corresponding *p-value*=0.101>0.05).

Another way to assess the goodness of fit is to compare the percentage of subjects correctly classified by the model with the percentage of subjects correctly classified by chance. From the results of the cross-tabulation between observed weight categories and predicted response weight categories (Table 1), we observe that the model predicts correctly 52.2% of the Underweight category, 82.7% of the Normal Weight category, 66.7% of the Overweight category and 52.6% of the Obese category.

In general the model classifies correctly 68.5% of the subjects. By comparison, the percentage of subjects correctly classified by chance is only 29%, so this model improves the classification.

The significance test of the model, based on the statistic G (equation 6), leads to the rejection of the null hypothesis that all coefficients of the variables in the model are zero (G = 234.173, *p-value*<0.05).

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

			Predicted Response Category				Total
			Under-weight	Normal weight	Over-weight	Obese	
Weight categories	Under-weight	Count	12	9	1	1	23
		% within Weight categories	52.2%	39.1%	4.3%	4.3%	100,0%
	Normal weight	Count	3	43	6	0	52
		% within Weight categories	5.8%	82.7%	11.5%	.0%	100,0%
	Over-weight	Count	0	7	22	4	33
		% within Weight categories	0.0%	21.2%	66.7%	12.1%	100,0%
	Obese	Count	1	3	5	10	19
		% within Weight categories	5.3%	15.8%	26.3%	52.6%	100,0%
Total		Count	16	62	34	15	127
		% within Weight categories	12.6%	48.8%	26.8%	11.8%	100.0%

The parallel lines test was chosen to make a judgement concerning the model's adequacy. The null hypothesis was that the corresponding regression coefficients were equal across all levels of the outcome variable. The chi-square test result (*p-value*>0.05) in this reduced model indicates that there are no significant differences for the corresponding regression coefficients across the response categories, suggesting that the model assumption of parallel lines is not violated.

The actual model proved to be the best when compared with other reduced models, since the Likelihood ratio test rejects the null hypothesis that the coefficients of the excluded variables are zero (p-value<0.05).

The significance test of the model, based on the statistic G (equation 6), leads to the rejection of the null hypothesis that all coefficients of the variables in the model are zero (G = 234.173, *p-value*<0.05).

Several R^2-like statistics can be used to measure the strength of the association between the dependent variable and the predictor variables. However they are not as useful as the statistic in regression, since their interpretation is not straightforward.

One of the commonly used statistics is the Cox and Snell R^2,

$$R^2_{CS} = 1 - \left(\frac{L\left(B^0\right)}{L\left(\hat{B}\right)} \right)^{\frac{2}{n}}$$

where $L(\hat{B})$ is the log-likelihood function for the model with the estimated parameters, $L(B^0)$ is the log-likelihood with just the thresholds and n is the number of cases. In this case, the value of the pseudo R^2 statistic is 0.539, the highest value we found among the considered set of significant and well-fitting models.

At a significance level of 5%, the most significant results associated with an increased risk of obesity were (Table 2): phenotype Hp2-1 of haptoglobin – this gives a higher chance of obesity compared with the reference category Hp2-2 ($\overline{OR} = 11.54$); phenotype AA of acid phosphatase – compared with the reference category (CC) it gives an almost 34 times greater probability of higher weight categories ($\overline{OR} = 33.79$); the estimated odds ratio for age, 1.39, indicates that for every increase of ten years in age, the risk of belonging to a higher weight category increases 7.613 times.

The interaction between phenotype Hp2-1 and phenotype AC of acid phosphatase is associated with a decreased risk of obesity ($\overline{OR} = 0.032$). The interaction between phenotype AA of acid phosphatase and the activity of this enzyme is associated with a decreased risk of obesity ($\overline{OR} = 0.95$).

Table 2. Parameter estimates, odds ratios and confidence interval estimates (Link function: Cauchit)

						95% Confidence Interval			95% Confidence Interval	
	Estimate	Std. error	Wald	df	p-value	Lower Bound	Upper Bound	$\overline{OR}$	Lower Bound	Upper Bound
Threshold										
[under-weight]	7.319	2.774	6.962	1	0.008	1.882	12.756			
[normal weight]	12.127	3.316	13.378	1	0.000	5.629	18.626			
[excess weight]	15.505	3.727	17.312	1	0.000	8.201	22.809			
Location										
[Hp1-1]	1.737	2.176	0.637	1	0.425	-2.527	6.000	5.013	0.136	18.024
[Hp2-1]	3.908	1.676	5.440	1	0.020	0.624	7.192	11.542	2.102	21.741
[Hp2-2]	(a)	.	.		.	.	.	(a)	(a)	(a)
[AA]	11.043	4.279	6.660	1	0.010	2.656	19.430	33.788	7.725	60.095
[AB]	3.249	2.671	1.479	1	0.224	-1.986	8.484	9.521	0.174	25.777
[AC]	7.082	4.513	2.463	1	0.117	-1.763	15.928	21.397	0.197	49.105
[BB]	0.901	2.690	0.112	1	0.738	-4.372	6.174	2.751	0.077	18.565
[BC]	0.034	3.975	0.000	1	0.993	-7.757	7.826	1.045	0.043	23.718
[CC]	(a)	.	.	0	.	.	.	(a)	(a)	(a)
ACP1 Activity	0.019	0.010	3.605	1	0.058	-0.001	0.040	1.025	0.999	1.052
Age	0.262	0.078	11.378	1	0.001	0.110	0.414	1.390	1.150	1.666
[Hp1-1] * [AA]	1.428	3.164	0.204	1	0.652	-4.774	7.631	4.144	0.070	23.109
[Hp1-1] * [AB]	-1.646	2.507	0.431	1	0.511	-6.559	3.267	0.210	0.051	9.575
[Hp1-1] * [AC]	-6.070	6.137	0.978	1	0.323	-18.098	5.958	0.055	0.018	17.891
[Hp1-1] * [BB]	-5.039	3.069	2.695	1	0.101	-11.054	0.977	0.067	0.030	2.941
[Hp1-1] * [BC]	1.770	2.680	0.436	1	0.509	-3.483	7.024	5.110	0.098	21.214
[Hp1-1] * [CC]	(a)	.	.	0	.	.	.	(a)	(a)	(a)
[Hp2-1] * [AA]	-4.433	2.392	3.434	1	0.064	-9.121	0.255	0.076	0.036	1.379
[Hp2-1] * [AB]	-0.746	1.646	0.205	1	0.650	-3.972	2.480	0.420	0.085	7.197
[Hp2-1] * [AC]	-10.246	3.612	8.046	1	0.005	-17.326	-3.166	0.032	0.019	0.108
[Hp2-1] * [BB]	-2.879	1.891	2.317	1	0.128	-6.586	0.828	0.119	0.050	2.573
[Hp2-1] * [BC]	-2.619	2.287	1.312	1	0.252	-7.102	1.863	0.131	0.047	5.377
[Hp2-1] * [CC]	(a)	.	.		.	.	.	(a)	(a)	(a)
[Hp2-2] * [AA]	(a)	.	.		.	.	.	(a)	(a)	(a)
[Hp2-2] * [AB]	(a)	.	.		.	.	.	(a)	(a)	(a)
[Hp2-2] *	(a)	.	.		.	.	.	(a)	(a)	(a)

[AC]									
[Hp2-2] * [BB]	(a)	.	.	.	.	.	(a)	(a)	(a)
[Hp2-2] * [BC]	(a)	.	.	.	.	.	(a)	(a)	(a)
[Hp2-2] * [CC]	(a)	.	.	.	.	.	(a)	(a)	(a)
[AA] * ACP1 activity	-0.037	0.016	5.182	1	0.023	-0.069	-0.005	0.954	0.916	0.993
[AB]* ACP1 activity	-0.014	0.010	1.818	1	0.177	-0.034	0.006	0.983	0.958	1.008
[AC] * ACP1 activity	-0.003	0.013	0.046	1	0.830	-0.029	0.023	0.996	0.964	1.030
[BB]* ACP1 activity	-0.001	0.011	0.013	1	0.911	-0.022	0.020	0.998	0.972	1.025
[BC]* ACP1 activity	0.001	0.013	0.002	1	0.964	-0.024	0.025	1.001	0.970	1.033
[CC] * ACP1 activity	(a)	.	.	0	.	.	.	(a)	(a)	(a)
ACP1 activity * age	0.000	0.000	3.802	1	0.051	-0.001	1.8E-6	1.000	0.999	1.000
[Female] * age	-0.052	0.036	2.093	1	0.148	-0.121	0.018	0.936	0.857	1.024
[Male] * age	(a)	.	.	0	.	.	.	(a)	(a)	(a)

(a) estimates not evaluated because they are redundant.

The odds ratio of outcome Y=i versus outcome Y=3 (reference category) for regressor values of **x**=**a** versus **x**=**b**, is

$$OR_i(\mathbf{a},\mathbf{b}) = \frac{P(Y=i|\mathbf{x}=\mathbf{a})/P(Y=3|\mathbf{x}=\mathbf{a})}{P(Y=i|\mathbf{x}=\mathbf{b})/P(Y=3|\mathbf{x}=\mathbf{b})} \quad , i=0, 1, 2,$$

where **x** is the dummy variable vector, taking 1 in the position corresponding to the presence of a specific category of the specific independent categorical variable.

For example, the odds ratio of outcome Y=i versus Y=3 for haptoglobin phenotype Hp1-1 versus Hp2-1, is

$$\overline{OR}_0 = (\mathbf{Hp1-1}, \mathbf{Hp2-1}) = \frac{P(Y=0\,|\,\mathbf{x}=(1,0,0))/P(Y=3\,|\,\mathbf{x}=(1,0,0))}{P(Y=0\,|\,\mathbf{x}=(0,1,0))/P(Y=3\,|\,\mathbf{x}=(0,1,0))} = 1.049.$$

The result indicates that there is no difference between the two phenotypes (Hp1-1, Hp2-1) as regards predisposition to obesity.

The other odds ratios are obtained similarly, and the results are shown in Table 3.

Table 3. Odds ratio estimates, $\overline{OR_i}(\mathbf{a},\mathbf{b})$.
*Reference category of the outcome variable is Y=3

Categories of independent variables (**a**)	*Out-come Y	Reference categories of independent variables (**b**)				
		Hp2-1	AB	AC	BB	BC
Hp1-1	0	1.049				
	1	1.049				
	2	1.015				
AA	0		0.072	0.133	0.062	0.069
	1		0.628	0.745	0.618	0.616
	2		0.757	0.885	0.748	0.746
AB	0		**	1.857	0.963	0.956
	1		**	1.187	0.985	0.981
	2		**	1.168	0.988	0.985
AC	0		**	**	0.519	0.515
	1		**	**	0.829	0.827
	2		**	**	0.846	0.843
BB	0		**	**	**	0.993
	1		**	**	**	0.997
	2		**	**	**	0.846

** Estimates not evaluated because they are redundant.

Examining Table 3, we find that:

- The phenotype 1-1 of haptoglobin carries a small increase in risk of obesity relative to phenotype Hp2-1;
- Regarding the acid phosphatase phenotypes, we find the following:

1. AA is a protecting factor against obesity relative to the other phenotypes AB, AC, BB, and BC;
2. AB is a risk factor for obesity relative to phenotype AC;
3. AC is a protecting factor against obesity relative to the remaining phenotypes BB and BC;
4. Finally, BB is a protecting factor against obesity relative to phenotype CC.

4. Conclusions

These results show three factors which are associated with an increased risk of obesity, namely: phenotype Hp2-1 of haptoglobin carries an almost 12 times higher probability of belonging to higher weight categories than phenotype Hp2-2 (reference category); phenotype AA of acid phosphatase carries an almost 34 times higher probability of belonging to higher weight categories than phenotype CC (reference category); for every increase of ten years in age, the risk of belonging to a higher weight category increases 7.613 times.

The interaction between phenotype Hp2-1 and phenotype AC of acid phosphatase is a protecting factor against obesity ($\overline{OR} = 0.032$) as is also the interaction between phenotype AA and the activity of this enzyme ($\overline{OR} = 0.954$). This last result suggests a question: why is phenotype AA a risk factor for obesity when isolated (main effect) but a protecting factor when associated with acid phosphatase activity?

Considering the results given in Table 5 (namely the 95% C.I.) and the value of the Cox and Snell R^2 (0.539), this model it is not a good model for prediction.

To conclude, this work has made it possible to identify some potential risk factors and relationships between them for obesity, i.e., knowing the genetic and demographic characteristics studied here, it will be possible to develop preventive measures, adapted to each particular case.

It seems important for further studies to be carried out, in order to confirm some of the results and to explore the potential relationship of excess weight and obesity with other variables.

REFERENCES

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

Giblett E.R. (1969). Genetic Markers in Human Blood. Blackwell Scientific.

Guerra A., Rego C., Castro E.B.M., Sinde S., Silva D., Rodrigues P., Crespo M.E., Albergaria F., Bicho M. (2000): Fosfatase acida, polimorfismo genético e factores de risco cardiovascular numa população pediátrica. Revista Portuguesa de Cardiologia 19: 679-91.

Gutteridge J.M. (1987): The antioxidant activity of haptoglobin towards haemoglobin-stimulated lipid peroxidation. Biochim. Biophys. Acta 917: 219-223.

Hosmer D.W., Lemeshow, S. (2000): Applied Logistic Regression. 2nd ed. John Wiley & Sons.
McCullagh,P., Nelder, J.A. (1989): Generalized Linear Models, 2nd ed., Chapman and Hall.
SPSS 16.0. http://www.spss.com/spss/whats_new_base.htm.