

Konrad Oexle

## Biochemical data in ornithine transcarbamylase deficiency (OTCD) carrier risk estimation: logistic discrimination and combination with genetic information

Received: 9 September 2005 / Accepted: 4 November 2005 / Published online: 2 February 2006  
© The Japan Society of Human Genetics and Springer-Verlag 2006

**Abstract** One-fifth of the gene mutations causing ornithine transcarbamylase deficiency cannot be detected. In such cases carrier risk estimation must refer to biochemical data—such as increased plasma glutamine concentration or increased orotidine excretion after allopurinol load—although these parameters do not yield a definite diagnosis. Here, I derive odds for carrier risk estimation from published data, i.e. from mean and standard deviation of glutamine concentrations in carriers and noncarriers, assuming normal distributions, and from allopurinol test results in individual carriers and noncarriers using logistic regression. I show how such biochemical information may be combined with genetic information, thus demonstrating the usefulness of biochemical data. The necessity to assess individual results in larger proband groups and to consider possible correlations between different parameters is indicated.

**Keywords** Ornithine transcarbamylase deficiency · Logistic regression · Odds · Glutamine · Allopurinol test · Bayes' theorem · Carrier risk estimation

### Introduction

Enzyme deficiencies along the pathway from ammonia to urea constitute the “urea cycle disorders” and result in hyperammonemic encephalopathy. Ammonia that is transformed to carbamylphosphate normally enters the urea cycle via ornithine transcarbamylase (OTC). In the case of OTC deficiency (OTCD), carbamylphosphate

drains into the pyrimidine synthesis pathway and enhances production of orotate and orotidine. Orotic aciduria therefore sets OTCD apart from other urea cycle disorders (Bachmann and Colombo 1980).

OTCD is an X-chromosomal disorder with an estimated incidence of 1:14,000 (Brusilow and Horwich 2001). In a typical male patient, encephalopathy develops soon after birth when placental clearance of ammonia has ended. Female carriers also have different levels of reduced enzyme activity and are endangered by metabolic stress. Usually, however, the medical history of carriers is unremarkable (Maestri et al. 1998), and the excretion of orotate is only slightly increased (Bachmann and Colombo 1980).

Since male OTCD patients do not reproduce, Haldane's (1935) equilibrium consideration predicts that about one-third of all OTCD gene mutations occur de novo. However, the mutation rate is much higher in spermatogenesis than in oogenesis (Tuchman et al. 1995). Thus, the mothers of four-fifths rather than two-thirds of all male patients are carriers, while in four-fifths of heterozygous females the mutation occurred de novo (McCullough et al. 2000). The rate of detection of OTCD mutations is limited to 80% (Tuchman et al. 2002). Thus, there is a considerable fraction of OTCD families where carriership in females is possible but uncertain.

Oxypurinol ribonucleotide, an allopurinol metabolite, enhances the excretion of orotate and orotidine by inhibiting orotidine decarboxylase. Therefore, the “allopurinol test” has been proposed as a tool for the detection of OTCD carriers (Hauser et al. 1990; Maestri et al. 1998). Recent analyses have shown, however, that the reliability of this test is limited (Bonham et al. 1999; Oexle et al. 2002; Grunewald et al. 2004). The rate of type II errors (carriership despite a negative test) in mothers of male patients may be as high as 23–34% (Oexle et al. 2002; Grunewald et al. 2004). Other biochemical parameters, such as serum glutamine concentration, also do not definitively distinguish carriers from noncarriers. However, biochemical data can impart useful information. Using logistic regression and

K. Oexle  
Institute of Clinical Genetics,  
Medical Faculty “Carl Gustav Carus”,  
Dresden University of Technology,  
Fetscherstrasse 74, 01307 Dresden, Germany  
E-mail: konrad.oexle@tu-dresden.de  
Tel.: +49-351-4585199  
Fax: +49-351-4584316

Bayesian analysis I show here how these data can be used and how they can be combined with genetic information in estimating carrier risk in OTCD.

### Combination of information

The joint probability  $P(C,D)$  of the two states C (e.g., “carrier”) and D is the product of the probability of one state times the conditional probability of the other. Thus,  $P(C,D) = P(D) P(C|D) = P(C) P(D|C)$  and, analogously,  $P(\bar{C}, D) = P(D) P(\bar{C}|D) = P(\bar{C}) P(D|\bar{C})$ , where the state  $\bar{C}$  (noncarrier) is the complement of state C. Dividing  $P(C,D)$  by  $P(\bar{C}, D)$  and taking the logarithm yields

$$\log\left(\frac{P(C|D)}{P(\bar{C}|D)}\right) = \log\left(\frac{P(D|C)}{P(D|\bar{C})}\right) + \log\left(\frac{P(C)}{P(\bar{C})}\right), \quad (1)$$

where  $P(C) = 1 - P(\bar{C})$  is the *prior* probability of state C,  $P(C|D) = 1 - P(\bar{C}|D)$  is the *posterior* probability of state C given observed data D, and  $\log(P(D|C)/P(D|\bar{C}))$  is the logarithm of odds (*lod*).

The mutually independent data values  $D_1$  and  $D_2$  may influence the probability of C. After taking information  $D_1$  into consideration, the prior probabilities  $P(C)$  and  $P(\bar{C})$  must be replaced by the posterior probabilities  $P(C|D_1)$  and  $P(\bar{C}|D_1) = 1 - P(C|D_1)$ , respectively. Considering  $D_2$  in the next step then gives Eq. 1 as

$$\log\left(\frac{P(C|(D_1, D_2))}{P(\bar{C}|(D_1, D_2))}\right) = \log\left(\frac{P(D_2|C)}{P(D_2|\bar{C})}\right) + \log\left(\frac{P(D_1|C)}{P(D_1|\bar{C})}\right) + \log\left(\frac{P(C)}{P(\bar{C})}\right). \quad (2)$$

The *lods* of mutually independent data  $D_1, \dots, D_r$  may thus be added,

$$\log\left(\frac{P((D_1, \dots, D_r)|C)}{P((D_1, \dots, D_r)|\bar{C})}\right) = \sum_i^r \log\left(\frac{P(D_i|C)}{P(D_i|\bar{C})}\right). \quad (3)$$

### Logistic regression

Logistic regression may be applied to describe how a random variable,  $X$ , influences a binary alternative. Consider a proband who, depending on  $X$ , has the probabilities  $P(C|X)$  and  $P(\bar{C}|X) = 1 - P(C|X)$  of being a carrier or not, respectively. In logistic regression the “logit”, i.e. the natural logarithm of  $P(C|X)/P(\bar{C}|X)$ , is modelled as a linear function of  $X$ ,

$$\log_e\left(\frac{P(C|X)}{1 - P(C|X)}\right) = a_0 + a_1X, \quad (4)$$

where  $a_0$  and  $a_1$  are constants. Taking the natural exponential function on both sides results in  $P(C|X)$  as a function of  $X$ ,

$$P(C|X) = \frac{\exp(a_0 + a_1X)}{1 + \exp(a_0 + a_1X)}, \quad (5)$$

with an S-shaped graph between  $P(C|X) \rightarrow 0$  for small  $X$  and  $P(C|X) \rightarrow 1$  for large  $X$ . The coefficients  $a_0$  and  $a_1$  are estimated by fitting this curve to empirical data points derived from a group of probands whose  $P(C|X)$  are known, i.e. 1 for obligate carriers and 0 for noncarriers. The fitting procedure (maximum likelihood) makes an assumption regarding the error distribution (e.g. not depending on  $X$ ) and hinges on the fact that small errors are more probable than large ones. The best fit thus results in the highest joint probability (likelihood  $L$ ) of the data points with respect to the regression curve. Maximizing  $L$  with respect to  $a_0$  and  $a_1$ , i.e.  $\partial L/\partial a_0 = 0$  and  $\partial L/\partial a_1 = 0$ , yields the estimated values of  $a_0$  and  $a_1$ . For calculation (Newtonian–Raphson method), commercial software (e.g. release 8.1 SAS/STAT of SAS, Cary, NC) or free Internet programs (Pezzullo 2005) may be used.

Logistic regression yields the odds (i.e. the *lod*) since the prior probability of carriership  $P(C)$  in the proband group is indicated by the fraction of  $N_C$  of carriers among the  $N$  probands on which the regression is based. With Eq. 1 the *lod* is

$$\log_e\left(\frac{P(X|C)}{P(X|\bar{C})}\right) = \log_e\left(\frac{P(C|X)}{1 - P(C|X)}\right) - \log_e\left(\frac{N_C}{N - N_C}\right). \quad (6)$$

The last term is zero if half of the probands are carriers, i.e.

$$\log_e\left(\frac{N_C}{N - N_C}\right) = \log_e 1 = 0.$$

### Logistic regression of allopurinol test results

Only a few allopurinol test results of individual obligate OTCD carriers are available in the literature. Results of four obligate carriers and four individuals from the general population can be extracted from Sebesta et al. (1994). Peak orotidine excretions following 300 mg allopurinol were 14.6, 15.3, 23.1, and 36.4 mmol/(mol creatinine) in carriers (Table 3 in Sebesta et al. 1994) and 1.7, 3.9, 4.3, and 6.4 mmol/(mol creatinine) in noncarriers (Fig. 2 in Sebesta et al. 1994). The authors observed that their own results as well as the results of Hauser et al. (1990) were skewed slightly towards large values. They therefore proposed logarithmic transformation. Thus, using the natural logarithms of the allopurinol test results of Sebesta et al. (1994), logistic regression yields

$$\log_e\left(\frac{P(C|\log_e(x))}{1 - P(C|\log_e(x))}\right) = -107.75 + 47.54 \log_e(x), \quad (7)$$

where  $x$  is the peak orotidine excretion in mmol/(mol creatinine). The same linear approximation applies to the *lod* since  $N_C = N - N_C = 4$  (see Eq. 6).

Carriership and noncarriership have the same probability at  $x = 9.65$ ; ratios of 1:100 and 100:1 are reached at  $x = 8.76$  and  $x = 10.63$ , respectively. The steepness of this regression curve results from the small and clearly separated proband groups. This separation does not represent the real situation, however. Grunewald et al. (2004) observed a large overlap between carriers and noncarriers, with 2 of 22 carriers having negative results and 6 of 20 noncarriers having positive results. Unfortunately, these authors did not provide the individual test results.

### Deriving odds from normal distributions

Individual biochemical data, such as plasma glutamine concentrations, in sufficiently large groups of OTCD carriers are not available. Only mean and standard deviation (SD) are indicated in the literature. Nonetheless, the odds may be derived if an assumption regarding data distributions is made. Here, I assume normal distributions, i.e. the probability density function,

$$f(x) = \frac{1}{\sigma(\sqrt{2\pi})} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad (8)$$

in carriers and noncarriers with the expected means  $\mu_C$  and  $\mu_{\mathcal{C}}$  and variances  $\sigma_C^2$  and  $\sigma_{\mathcal{C}}^2$ , respectively. Using the probability density functions  $f_C(x)$  and  $f_{\mathcal{C}}(x)$  in carriers and noncarriers, respectively, the *lod* can be calculated as

$$\log_e\left(\frac{P(x|C)}{P(x|\mathcal{C})}\right) = \log_e\left(\frac{f_C(x)}{f_{\mathcal{C}}(x)}\right) = \log_e\left(\frac{\sigma_{\mathcal{C}}}{\sigma_C}\right) + \frac{(x-\mu_{\mathcal{C}})^2}{2\sigma_{\mathcal{C}}^2} - \frac{(x-\mu_C)^2}{2\sigma_C^2}. \quad (9)$$

The odds are 1 (and *lod* = 0) at position  $x_0$  where  $P(x_0|C) = P(x_0|\mathcal{C})$ , i.e. where  $x$  has the same probability density in carriers (C) and in noncarriers ( $\mathcal{C}$ ). Thus, Eq. 9 results in a quadratic equation with the solution

$$x_0 = A \pm \sqrt{\frac{A^2 - \frac{\sigma_C^2\mu_{\mathcal{C}}^2 - \sigma_{\mathcal{C}}^2\mu_C^2 + 2\sigma_C^2\sigma_{\mathcal{C}}^2 \log_e(\sigma_{\mathcal{C}}/\sigma_C)}{\sigma_C^2 - \sigma_{\mathcal{C}}^2}}{\sigma_C^2 - \sigma_{\mathcal{C}}^2}}, \quad (10a)$$

$$A = \frac{\sigma_C^2\mu_{\mathcal{C}} - \sigma_{\mathcal{C}}^2\mu_C}{\sigma_C^2 - \sigma_{\mathcal{C}}^2},$$

$$x_0 = \frac{\mu_C + \mu_{\mathcal{C}}}{2} \quad \text{if } \sigma_C = \sigma_{\mathcal{C}}. \quad (10b)$$

The first derivative at  $x_0$  yields the linear approximation in the vicinity of  $x_0$ ,

$$\log_e\left(\frac{P(x|C)}{P(x|\mathcal{C})}\right) = (x-x_0) \frac{\partial}{\partial x} \log_e\left(\frac{f_C(x)}{f_{\mathcal{C}}(x)}\right)_{x=x_0} = (x-x_0) \left(\frac{x_0 - \mu_{\mathcal{C}}}{\sigma_{\mathcal{C}}^2} - \frac{x_0 - \mu_C}{\sigma_C^2}\right) = b_0 + b_1x, \quad (11)$$

where  $b_0$  and  $b_1$  are constants.

### Odds of the plasma glutamine concentration

Consider, for instance, the plasma glutamine concentrations in OTCD carriers (mean  $\pm$  SD = 702  $\pm$  167  $\mu\text{mol/l}$ ) and noncarriers (568  $\pm$  84.4  $\mu\text{mol/l}$ ) as determined by Maestri et al. (1998). With these estimations on the  $\mu$  and  $\sigma$  values of the assumed normal distributions and their probability density functions  $f_C(x)$  and  $f_{\mathcal{C}}(x)$ , the linear approximation above (see Eqs. 10a, 10b, 11) yields

$$\log_e\left(\frac{P(x|C)}{P(x|\mathcal{C})}\right) = b_0 + b_1x = -10.20 + 0.015x, \quad (12)$$

in the vicinity of a plasma glutamine concentration of  $x_0 = 668 \mu\text{mol/l}$ . The odds  $P(x|C)/P(x|\mathcal{C})$  are identical to the quotient  $P(C|x)/P(\mathcal{C}|x) = P(C|x)/(1 - P(C|x))$  if the prior probability is 50%, i.e. if  $P(C) = 0.5 = P(\mathcal{C})$  (see Eq. 3). The posterior probability  $P(C|x)$  is then given as  $P(C|x) = (\text{odds})/(1 + \text{odds})$ . Figure 1 displays the probability density functions  $f_C(x)$  and  $f_{\mathcal{C}}(x)$ , and, assuming  $P(C) = 0.5$ , the posterior probability  $P(C|x)$  as derived from Eq. 9 and as approximated in Eq. 12. The approximation deviates considerably for concentrations outside the interval 668  $\pm$  70  $\mu\text{mol/l}$ .

### Associated random variables

Biochemical variables, such as glutamine and ammonia concentrations, for instance, are not always independent. If two associated random variables  $X$  and  $Y$  are used in carrier risk calculation, logistic regression must involve in both of them, thus resulting in

$$\log_e\left(\frac{P(C|(X, Y))}{1 - P(C|(X, Y))}\right) = d_0 + d_1X + d_2Y. \quad (13)$$

Equation 13 is identical to the sum of the individual regression of the two variables if both variables are independent and the regression is performed on equal numbers of carriers and noncarriers, i.e.  $P(C) = 0.5 = P(\mathcal{C})$ , as follows immediately from the considerations above (see Eqs. 2, 4).

### Example

Consider the case of a woman who lost a son from OTCD. Molecular genetic examination of this son's DNA was not performed. Analysis of the woman's DNA does not reveal an OTCD gene mutation. She has a second son who is healthy. Her plasma glutamine concentration is 600  $\mu\text{mol/l}$ , i.e. below the value where the likelihood ratio (odds) is 1 (see Eq. 12). What is her risk of being a carrier ("C")? The prior probability  $P(C)$  in mothers of male OTCD patients is 0.8 (McCullough et al. 2000). In the present example, this probability is modified by three independent pieces of information, i.e.

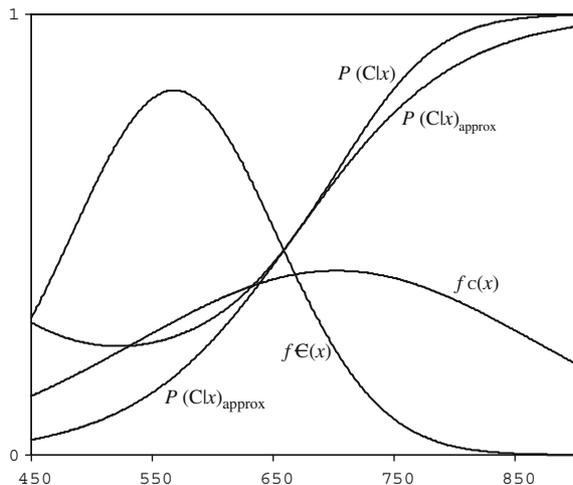
by (1) the nondetection “M” of a mutation in the mother’s DNA, (2) the existence “U” of an unaffected son, and (3) the plasma glutamine concentration  $x$ . Independence means that the detectability of a carrier’s mutation, her plasma glutamine level, and the likelihood of her unaffected son do not influence each other. Equation 2 therefore yields

$$\log_e \left( \frac{P(C|M, U, x)}{P(\bar{C}|M, U, x)} \right) = \log_e \left( \frac{P(M|C)}{P(M|\bar{C})} \right) + \log_e \left( \frac{P(U|C)}{P(U|\bar{C})} \right) + \log_e \left( \frac{P(x|C)}{P(x|\bar{C})} \right) + \log_e \left( \frac{P(C)}{P(\bar{C})} \right). \quad (14)$$

where  $P(M|C)$  is 0.2, i.e. 1 – the mutation detection rate in OTCD as indicated by Tuchman et al. (2002).  $P(M|\bar{C})$  is 1 since noncarriers do not have OTCD mutations.  $P(U|C)$ , the probability that a carrier of the X-chromosomal disorder of the unaffected son, is 0.5, while  $P(U|\bar{C})$  is almost 1 since the incidence of OTCD is small in the general population. Equation 12 yields  $\log_e (P(x|C)/P(x|\bar{C})) = -10.20 + 0.015 \times 600 = -1.20$ . With Eq. 14

$$\begin{aligned} \log_e \left( \frac{P(C|M, U, x)}{1 - P(C|M, U, x)} \right) &= \log_e \left( \frac{0.2}{1} \right) + \log_e \left( \frac{0.5}{1} \right) \\ &\quad - 1.20 + \log_e \left( \frac{0.8}{0.2} \right) \\ &= -2.12, \end{aligned} \quad (15)$$

the posterior probability can thus be calculated as  $P(C|M, U, 600) = 0.11$ . Hence, by including the genetic



**Fig. 1** Probability density functions  $f_C(x)$  and  $f_{\bar{C}}(x)$  of plasma glutamine concentration in ornithine transcarbamylase deficiency (OTCD) carriers and noncarriers, respectively, assuming normal distributions (mean  $\pm$  SD values) as determined by Maestri et al. (1998). Posterior carrier probability  $P(C|x)$  as derived from Eqs. 3 to 9 assuming a prior probability of 50%. Approximation  $P(C|x)_{approx}$  of this posterior probability as determined for the interval comprising the intersection of  $f_C(x)$  and  $f_{\bar{C}}(x)$  at  $x_0 = 668 \mu\text{mol/l}$  (see Eq. 12)

and biochemical data the carrier risk has been reduced from 80% to 11%.

## Discussion

Unambiguous diagnosis (or exclusion) of OTCD carrier-ship cannot be derived from biochemical data since carriers and noncarriers overlap. This is true for plasma glutamine concentration (Fig. 1) as well as orotidine and orotate excretion after allopurinol load (Grunewald et al. 2004). Nonetheless, the biochemical data may be used in OTCD carrier risk estimation as I have shown here. Similar logistic regression analyses of biochemical data have been performed before in other disorders, e.g. logistic regression of the logarithm of creatine kinase activity in carriers and noncarriers of Duchenne muscular dystrophy (Percy et al. 1981, 1987).

The concrete quantitative results provided here should be considered as examples only, however. The proband groups used for logistic regression of the allopurinol test data were too small and did not sufficiently represent the overlap between carriers and noncarriers. Therefore, Eq. 7 should not be used in carrier risk estimation. It is meant as a suggestion of how one could analyse representative proband groups such as those tested by Grunewald et al. (2004) if individual data were available.

The odds derived for plasma glutamine concentration (Eq. 12) may be more realistic. However, the assumption that this concentration has normal distributions in carriers and noncarriers of OTCD needs to be validated. Therefore, if the glutamine concentration in a possible carrier is unusually small, for instance, its use according to Eqs. (12) and (14) may result in an incorrectly low carrier probability. Logistic regression of glutamine concentrations in individual probands would provide a more reliable basis. Even then, however, the caveat of Percy et al. (1987) would apply; this states that in order to prevent misclassification of carriers as normal individuals, no logistic carrier probability should be lower than  $1/N_C$  (where  $N_C$  is the number of obligate carriers on which the logistic regression was based).

Frequently, direct mutation detection or indirect marker analysis may indicate a carrier status sufficiently well. However, such analyses are not always definite. Indeed, only 80% of all OTCD gene mutations can be detected (Tuchman et al. 2002). Nonetheless, in order to achieve adequate carrier risk estimation, nondefinite information derived from molecular genetic analyses must be taken into account. How this may be done has been shown in the example above where the nondetection of an OTCD gene mutation in a possible carrier has been combined with pedigree information and biochemical data. Indirect genetic analysis using DNA markers produces odds that depend on the recombination probability between marker and gene. These odds could also be combined with other odds as shown in Eq. 14.

Pieces of information are not necessarily independent of each other. This certainly applies to biochemical data and, indeed, allopurinol test results and plasma glutamine concentration in OTCD carriers may be correlated. To adequately account for different variables and their possible associations, multiple logistic regressions should be performed (Percy et al. 1981; Pezzullo 2005). For that purpose, all variables of interest have to be measured in sufficiently large groups of individual carriers and noncarriers.

---

## References

- Bachmann C, Colombo JP (1980) Diagnostic value of orotic acid excretion in heritable disorders of the urea cycle and in hyperammonemia due to organic acidurias. *Eur J Pediatr* 134:109–113
- Bonham JR, Guthrie P, Downing M, Allen JC, Tanner MS, Sharrard M, Rittey C, Land JM, Fensom A, O'Neill D, Duley JA, Fairbanks LD (1999) The allopurinol load test lacks specificity for primary urea cycle defects but may indicate unrecognized mitochondrial disease. *J Inher Metab Dis* 22:174–184
- Brusilow SW, Horwich AL (2001) Urea cycle enzymes. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic and molecular bases of inherited disease*, 8th edn. McGraw-Hill, New York, pp 1909–1963
- Grunewald S, Fairbanks L, Genet S, Cranston T, Husing J, Leonard JV, Champion MP (2004) How reliable is the allopurinol load in detecting carriers for ornithine transcarbamylase deficiency? *J Inher Metab Dis* 27:179–186
- Haldane JBS (1935) The rate of spontaneous mutation of a human gene. *J Genet* 31:317–326
- Hauser ER, Finkelstein JE, Valle D, Brusilow SW (1990) Allopurinol-induced orotidinuria. A test for mutations at the ornithine carbamoyltransferase locus in women. *N Engl J Med* 322:1641–1645
- Maestri NE, Lord C, Glynn M, Bale A, Brusilow SW (1998) The phenotype of ostensibly healthy women who are carriers for ornithine transcarbamylase deficiency. *Medicine* 77:389–397
- McCullough BA, Yudkoff M, Batshaw ML, Wilson JM, Raper SE, Tuchman M (2000) Genotype spectrum of ornithine transcarbamylase deficiency: correlation with the clinical and biochemical phenotype. *Am J Med Genet* 93:313–319
- Oexle K, Bonafé L, Steinmann B (2002) Remark on utility and error rates of the allopurinol test in detecting mild ornithine transcarbamylase deficiency. *Mol Genet Metab* 76:71–75
- Percy ME, Andrews DF, Thompson MW (1981) Duchenne muscular dystrophy carrier detection using logistic discrimination: serum creatine kinase and hemopexin in combination. *Am J Med Genet* 8:397–409
- Percy ME, Andrews DF, Brasher PMA, Rusk ACM (1987) Making most of multiple measurements in estimating carrier probability in Duchenne muscular dystrophy: the Bayesian incorporation of repeated measurements using logistic discrimination. *Am J Med Genet* 26:851–861
- Pezzullo JC (2005) Logistic regression. <http://www.members.aol.com/johnp71/logistic.html>
- Sebesta I, Fairbanks LD, Davies PM, Simmonds HA, Leonard JV (1994) The allopurinol loading test for identification of carriers for ornithine carbamoyl transferase deficiency: studies in a healthy control population and females at risk. *Clin Chim Acta* 224:45–54
- Tuchman M, Matsuda I, Munnich A, Malcolm S, Strautnieks S, Briede T (1995) Proportions of spontaneous mutations in males and females with ornithine transcarbamylase deficiency. *Am J Med Genet* 55:67–70
- Tuchman M, Jaleel N, Morizono H, Sheehy L, Lynch MG (2002) Mutations and polymorphisms in the human ornithine transcarbamylase gene. *Hum Mutat* 19:93–107