

responsible for the Bohr effect in the Hb A molecule. The main aspect of this investigation consists of the role of $\beta 146\text{His}$, an amino acid residue previously shown to contribute significantly to the alkaline Bohr effect in Hb A. We have found that, in the absence of phosphate ions and at low chloride ion concentrations, the pK value of $\beta 146\text{His}$ does not change significantly upon ligation; thus, under these experimental conditions, the contribution of $\beta 146\text{His}$ to the alkaline Bohr effect is negligible. We have concluded that the molecular mechanism of the Bohr effect is not only the result of the quaternary and tertiary structure changes brought about in the Hb A molecule upon ligation, but it also involves specific interactions of one or several oxygen-linked groups with the solvent molecules or ions.

For the case of the Hb S molecule, we have found that three histidine residues in CO-Hb S and six histidine residues in deoxy-Hb S have pK values and proton intrinsic chemical shifts which are different from the corresponding ones in Hb A. These findings indicate that the surface conformation of the Hb S molecule is different from that of the Hb A molecule.

The molecular mechanism of the polymerization process of Hb S has been investigated by measuring the longitudinal relaxation rates (T_1^{-1}) of the C2 protons of the eleven observable histidine residues in deoxy-Hb S and deoxy-Hb A as a function of hemoglobin concentration and temperature. We have found that, under certain experimental conditions, several surface histidine C2 protons have increased T_1^{-1} values in deoxy-Hb S as compared to the corresponding ones in deoxy-Hb A. When the Hb S solution approaches gelation these differences in T_1^{-1} values between deoxy-Hb S and deoxy-Hb A are enhanced. This result suggests that the regions on the surface of the Hb S molecule on which these histidine residues are situated are involved in the early stages of the polymerization process. We propose that the increased T_1^{-1} values of several histidine resonances in deoxy-Hb S result from the formation in the Hb S solution of small pregelation aggregates prior to the main polymerization. The histidine residues for which the differences in T_1^{-1} values are the largest are viewed as being located at or near the "contact" areas between the Hb S molecules within the pregelation aggregate.

BIOSTATISTICS

A MULTIVARIATE ANALYSIS OF THE INCIDENCE OF ELEVATED BLOOD LEAD LEVEL IN CHILDREN

Order No. 8015388

AL-RAWI, ZEYAD RASHAD, PH.D. *University of Pittsburgh*, 1979. 170pp.

In the presented dissertation a major effort has been made to determine the environmental and socio-economic variables that increase the probability of a child between the ages of one and six having an elevated blood lead level (EBL). It has been shown, given that environmental lead sources exist in areas, a child's socio-economic variables will influence the probability of ingestion of abnormal quantities of lead. Eight socio-economic variables have been considered which have been found to have some association with the dependent variable - the Proportion of EBL. These variables are: X_1 = The proportion of house units built before 1940; X_2 = The proportion of population below poverty level; X_3 = The proportion of children ≤ 6 years old living with one parent families; X_4 = The proportion of population < 6 years old; X_5 = The proportion of blacks among children ≤ 6 years old; X_6 = The proportion of children 3 - 6 years old not attending pre-school; X_7 = The average family size for children ≤ 6 years old; and X_8 = The proportion of population ≥ 25 years old having ≤ 10 years schooling.

The 1970 census data were used to calculate these variables by census tract in Pittsburgh, Philadelphia, and Memphis. Analysis of variance studies were performed to see if the census tracts were homogeneous with regard to their mean blood lead level in the three cities. It was found that they were not.

A principal component analysis was performed to develop a model as a linear combination of these socio-economic variables which utilizes appropriate weights for these variables - variables with small variance which provide more information receive large weights and variables with large variance which provide little information receive small weights.

When the model developed from Pittsburgh data was applied to the data from two other cities, the percent of explained variation of EBL was 40.22% in Philadelphia and 32.77% in Memphis. For biological data, these percentages of explained variation would be considered large especially if this is done by using non-biological independent variables.

Other multivariate statistical procedures such as Fisher's discriminant function analysis and the multivariate logistic method, were performed to investigate and identify the census tracts with high risk of EBL as well as providing more information on the most important socio-economic variables associated with the EBL. The results of these analyses indicate that three socio-economic variables appear to make a significant contribution to the variation in the proportion of EBL in children. These variables are (X_1) the proportion of house units built before 1940, (X_2) the proportion of population below poverty level, and (X_3) the proportion of black children.

It is suggested that these three variables be given high priority whenever a screening program is planned for elevated blood lead level in children.

MEASUREMENT ERROR IN LOGISTIC REGRESSION BY DISCRIMINANT ANALYSIS WITH APPLICATIONS TO THE EPIDEMIOLOGY OF CORONARY HEART DISEASE

Order No. 8014607

BAWOL, RICHARD DAVID, PH.D. *University of California, Berkeley*, 1979. 131pp.

The effect of measurement error on logistic regression has been examined, primarily under normality assumptions when coefficients can be estimated by discriminant analysis and in the context of a prospective epidemiological study of coronary heart disease (CHD). The statistical problem is that of relating a dichotomous response to continuous variables when the continuous variables are measured with error. This occurs when the presence or absence of CHD during a follow-up period is related to baseline risk variables which are subject to intra-individual variation and errors in measurement procedures. Two general questions are studied in this dissertation: (1) How do observable relationships compare to true relationships? (2) How can true relationships be estimated from observations measured with error?

Literature on measurement error in regression analysis is reviewed to provide a source of ideas and a frame of reference. Of particular interest are the attenuation effects for regression coefficients, and maximum likelihood estimation for functional and structural relationships with measurement error.

The coefficient of reliability (variance of true/variance of observable) can be estimated from variance components or by the correlation between repeated measurements. The equivalence between these two types of estimators is established for each of three error models. When these estimators are applied to a diet reliability sample, differences between the error models appear to be inconsequential.

Attenuation effects of measurement error on logistic regression are formulated by relating discriminant coefficients for observables to those for true values. The primary effect is to attenuate a discriminant coefficient by a factor equal to the coefficient of reliability. A secondary effect which operates through correlated covariates is studied for a bivariate discriminant. Data from the Western Collaborative Group Study (WCGS) indicate the impact of measurement error in a prospective epidemiological study of CHD.

Two alternative models are also considered. The "flat slope syndrome" of negative correlation between measurement errors and true values may further attenuate a discriminant coefficient. A ln-linear dichotomous response model with measurement error again involves an attenuation factor equal to the coefficient of reliability.

Estimation and hypothesis testing for true univariate logistic regression by discriminant analysis of observations with error are investigated by deriving maximum likelihood estimators and their standard errors. The general idea is to adjust the discriminant coefficient to remove the effect of measurement error by estimating intra-individual variation from replicate observations. Loss of power due to measurement error and compensating sample size requirements are determined from normal theory. WCGS data are analyzed again, this time to estimate discriminant coefficients from replicate observations and to judge the effect of measurement error on the significance and standard error of estimates.

Finally, the results of this dissertation are summarized, with particular regard to epidemiological implications and directions for further study. Investigators are urged to report measurement reliability, to improve reliability by using averages of replicates, and to consider eventually adjusting risk associations for measurement error. Further study is needed to develop multivariate adjustment methods for general logistic regression with measurement error.