

National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 4-5, 2004

Hierarchical Aggregate Data Analysis
by
Nelia Sabater Ereno

For additional information, please contact:

Author's name:	Nelia Sabater Ereno
Designation:	Assistant Professor
Agency:	University of the Philippines-Cebu
Address:	c/o School of Statistics, University of the Philippines, Diliman Quezon City
Telefax:	(632) 920-5301 to 99 loc 6184, 6894, 6471
E-mail:	nelia.s.ereno@up.edu.ph

Hierarchical Aggregate Data Analysis

by

Nelia Sabater Ereno¹

ABSTRACT

The aggregate data study design defines a group as the unit of analysis. It is used when individual data is not available. Aggregate data are easily and relatively cheaply available as compared to individual data. In this paper, a hierarchical aggregate data model with spatial correlations between groups of interest will be formulated. The parameters involved in the model will then be estimated. The occurrence of biases such as misspecification due to confounding, effect modification and effect of within-area variability will be examined. Standardization, pure specification and cross-level biases will also be observed. Finally, simulations of the data from the Family Income and Expenditures Survey (FIES) will be utilized to validate and illustrate the hierarchical aggregate data model.

I. Introduction

Aggregate data analysis is a topic that has been developed greatly since the middle of the last decade. It offers significant improvement over ecological studies in the inference of individual behavior from aggregate data. It is utilized to infer individual-level relationships when the individual level data are not available. It has been widely used also in a variety of fields such as political science, marketing, sociology, epidemiology, public policy and quantitative history.

Like the small area and ecologic designs, aggregate data design uses group as the unit of analysis. It is concerned with the control of confounding and adjustment for spatial variation that are relevant also to small area analysis. Unlike in ecological study that utilizes group data, aggregate data study utilizes individual data and the models that are used are constructed from the individual level. In small area analysis, one focuses on the prediction of events while in aggregate data study, focus is on the estimation of exposure effects rather than on prediction of events.

Using group as the unit of analysis has many advantages such as: 1) measurement error may cause less bias in group-level than individual studies; 2) Group-level studies can estimate exposure effects that may be difficult to detect within any one group of individuals; 3) Disease rates and exposure often show more variation between rather than within areas; 4) Large groups such as those defined by cities, states, or countries are often available through public-use data sets. Hence, they can often be performed quickly and inexpensively; 5) Since large populations are considered, group studies provide a means of increasing statistical power.

¹ Nelia S. Ereno – Assistant Professor, University of the Philippines-Cebu

Generally, a random variable can have only one distribution. However, there are cases where it is easier to model situations in a hierarchy. This enables the complicated processes to be modeled by a sequence of relatively simple models that are placed in a hierarchy. Dealing with the hierarchy is not as much difficult as dealing with conditional and marginal distributions.

The most classical hierarchical model talks about a problem on an insect that lays a large number of eggs, each surviving with probability p . On the average, one would like to determine how many eggs will survive. If we assume that each egg's survival is independent, then we have Bernoulli trials. If we let X be the number of trials and Y be the number of eggs laid. Then,

$$X | Y \sim \text{Binomial}(Y, p)$$

$$Y \sim \text{Poisson}(\lambda)$$

is a hierarchical model.

Hierarchical aggregate data model is basically a combination of the aggregate data model and the hierarchical Bayesian disease mapping. We first discuss the aggregate data model.

II. The Aggregate Data Model

Suppose there are $k = 1, 2, \dots, K$ groups of interest, each consisting of $i = 1, \dots, n_k$ individuals. Outcomes will be stratified by confounders such as age and gender so that $n_k = \sum n_{ks}$ for strata $s = 1, \dots, S$. Let Y_{ksi} be a binary indicator, with $Y_{ksi} = 1$ for a case. Let us say,

$$Y_{ksi} = \begin{cases} 1 & \text{if individual } i \text{ in population } (k, s) \text{ develops the study} \\ & \text{disease} \\ 0 & \text{otherwise} \end{cases}$$

Assume that

$$Y_{ksi} | \alpha, x_{ksi}, v_k \sim \text{Bernoulli}(\exp\{\gamma_s + x_{ki}\beta + v_k\})$$

$$\text{where } \alpha = (\gamma_s, \beta)^T, \text{ where } \gamma_s = \gamma - \frac{1}{2} \alpha^T \Delta \alpha$$

$x_{ksi} = (1, Z_{ksi})$ where Z_{ksi} is the row vector of covariates for the i th member of the k th group in strata s .

v_k are spatially unstructured variation which accounts for the unknown or unmeasured disease risk factors.

β is the regression parameters to be estimated

Note that Δ is a normalization constant defined by

$$\Delta = \sum \exp \{y^t \theta + w^t \lambda + c(y)\}$$

The random effect v_k , $k=1, 2, \dots, K$, are independent and identically distributed with mean 0 and variance σ_v^2 . The relative rate parameter for the exposure, β , is assumed to be constant across strata. The parameters α and β are estimated using the estimating equation.

$$\sum_{k=1}^K D_k^t V_k^{-1} \mathbf{f}_k = \mathbf{0} .$$

where

$$D_k = \begin{pmatrix} \frac{\partial \mu_k}{\partial \beta^t} & 0 \\ \frac{\partial \sigma_k}{\partial \beta^t} & \frac{\partial \sigma_k}{\partial \alpha^t} \end{pmatrix} \quad \mathbf{f}_k = \begin{pmatrix} y_k - \mu_k \\ \mathbf{s}_k - \sigma_k \end{pmatrix}$$

$$V_k = \begin{pmatrix} V_{k11} & V_{k12} \\ V_{k21} & V_{k22} \end{pmatrix} = \begin{pmatrix} \text{var } y_k & \text{cov}(y_k, \mathbf{s}_k) \\ \text{cov}(\mathbf{s}_k, y_k) & \text{var } \mathbf{s}_k \end{pmatrix}$$

and where $\mathbf{s}_k^t = (s_{k12}, s_{k13}, \dots, s_{k23}, \dots)$, $s_{kij} = (y_{ki} - \mu_{ki})(y_{kj} - \mu_{kj})$ is the vector of covariances

The derivation of the estimating equation that solves the relative rate parameters α and β are found in the Appendix of this paper.

To model Y_{ks} which is simply the observed number of cases in strata s , group k , we have

$$Y_{ks} | \alpha, x_k, v_k \sim \text{Binomial}(n_{ks}, \exp\{Y_s\} p_k(\beta, x_k) \exp\{v_k\})$$

$$\text{where } x_k = (x_{k1}, \dots, x_{knk})^T$$

$$p_k(\beta, x_k) = \int \exp(x\beta) f_{ks}(x) dx.$$

If the distribution of the exposure is independent of stratum, we can write

$$p_k(\beta, x_k) = \int \exp(x\beta) f_k(x) dx.$$

As $n_{ks} \rightarrow \infty$ then we approximate the binomial by the Poisson model and sum over strata to obtain

$$Y_k | \alpha, x_k, v_k \sim \text{Poisson}(E_k p_k(\beta, x_k) \exp\{v_k\})$$

where $E_k = \sum n_{ks} \exp\{\gamma_s\}$

Note that E_k is the expected number of events.

The above results make the aggregate data model assumes only the Bernoulli model for the individual outcomes, with

$$E(Y_k|\alpha, v_k) = E_k p_k(\beta, x_k) \exp\{v_k\}$$

$$\text{where } p_k(\beta, x_k) = n_k^{-1} \sum \exp\{x_{ki} \beta\}.$$

III. Subsample Covariate Model

The method is extended to a more realistic scenario where only a subsample of covariate values is available by replacing the functions of the group covariates with the corresponding functions of the subsample data. For $m_k < n_k$, the sampled-covariate aggregate model is

$$E(Y_k|\alpha, v_k) = E_k p_k(\beta, x_k^m) \exp\{v_k\}$$

$$\text{where } x_k^m = (x_{k1}, \dots, x_{kmk})^T$$

$$p_k(\beta, x_{mk}) = m_k^{-1} \sum \exp\{x_{ki} \beta\}.$$

In the above model we considered the assumption that there is a spatial correlation between groups of interest since it is reasonable to assume that observations in neighboring areas may show similarities not attributable to measured covariates. These similarities may come from the patterns of cultural or ethnic affiliations or geographical characteristics.

IV. Bayesian Disease Mapping Model

Disease mapping is a process of describing geographical variation in disease incidence and mortality. A disease-mapping model may also include an ecological regression on potential risk factors for diseases. It is a Poisson random variate with

$$E(Y_k|\alpha, x_k, u_k, v_k) = E_k p_k(\beta, x_k) \exp\{u_k + v_k\}$$

where x_k is an area-level exposure value,
 $p_k(\beta, x_k) = \exp\{x_{ki} \beta\}$.
 u_k and v_k are the spatially clustered and unstructured variation, respectively.

The random effects are intended to account for unknown or unmeasured disease risk factors. Each u_k and v_k are assumed to be independent of one another, as well as independent of the exposure distribution.

The distribution of the random effect u_k is a Gaussian intrinsic autoregressive distribution. That is,

$$u_k | u_l \in \delta_k \sim N(u_k w_{k+}^{-1} \sigma_u^2)$$

where δ_k indicates the set of neighboring areas,
 w_{k+} is the number of neighbors,
 u_k is the mean of these neighbors.

Since we have a fully Bayesian model here, inference about the unknown parameters is of course based on the joint posterior distribution. The joint distribution, up to a normalizing constant, is analytically intractable, thus the Markov Chain Monte Carlo (MCMC) is used for estimation. MCMC relies on the setting up of a Markov chain designed so that its stationary distribution is the posterior distribution. It will simulate from the posterior distribution of the parameters given the data. The algorithms are the Metropolis-Hastings type which has the Gibb sampler as a special case.

V. The Complete Covariate Data Model

To combine the original aggregate data model and the Bayesian disease-mapping model, we have now a complete covariate data model. That is, we assume a Poisson model for the observed number of cases with

$$E(Y_k | \alpha, x_k, u_k, v_k) = E_k p_k(\beta, x_k) \exp\{u_k + v_k\}$$

$$\text{where } p_k(\beta, x_k) = n_k^{-1} \sum \exp\{x_{ki}\beta\}.$$

The Bernoulli random variates are not iid; therefore, the likelihood is not binomial.

To estimate the parameters of the hierarchical aggregate data model we use a slight variation on Gibbs sampling which is a special case of MCMC. We simulate values of β via a single-component Metropolis-Hastings algorithm.

VI. Biases

The aggregate data model estimates are subject to biases not present in estimates from individual-level observational studies of the same populations. These biases can invalidate the inference in group level studies.

Prentice and Sheppard (1995) showed that the linear and log-linear aggregate data models are not affected by the classical measurement errors in the covariates. On the other hand, Carroll (1997) showed the effect of measurement error on the Prentice and Sheppard “unweighted” aggregate data

analysis in the special case that an error-prone version (W_{ki}) of (X_{ki}) is measured along with the response (Y_{ki}); the true covariates (X_{ki}) are not observed. He found out that under regularity conditions and assuring that the number of observations per population is large, the effect of classical measurement errors in the probit model is reverse attenuation. That is, the classical measurement errors in the covariates can induce asymptotic bias away from the null in a probit aggregated model. In other words, attenuation bias never occurs in the probit aggregated model. It was said to be a surprising effect since it was the first known case under reasonable distributional assumptions that the effect of measurement error is reverse attenuation.

Measurement error is just one of the many sources of bias in ecological studies. Thus, an aggregate data analysis is a better alternative to ecological data analysis since unlike the latter, the former is not affected by classical measurement errors in the covariates (Guthrie and Sheppard, 2001).

Other ecological biases are the so-called pure specification bias and a cross-level bias. Bias also maybe due to within- and between-group confounding, effect modification, exposure misclassification, and lack of mutual standardization between the response and exposure.

In ecologic study, confounder misspecification is much more difficult to avoid than in individual-level study. For example, if in an individual-level study the main potential confounder is let us say the smoking history of an individual then in the corresponding ecologic study the analogous potential confounder will be the distribution of all smoking histories across all individuals within each region.

In typical ecologic studies, the summaries (such as the mean) available for each region may be inadequate to control confounding by the summarized covariate. It is true if nonlinearity or nonadditivity of effects is present at the individual level. Thus, we encounter a bias due to effect modification.

Standardization bias is common in ecological analysis because disease rates are usually aged-standardized while ecologic exposures and covariates are usually not. When adding summaries to regression, covariates need to be standardized using the same standard distribution as used for the outcome. Otherwise, addition of covariate to regression may even worsen bias. Exposure must also be standardized using the same standard distribution if bias is to be avoided. In general, if some but not all the variables in a regression have been standardized, or if different variables have been standardized to different distributions, severe bias may result.

The target of inference for both ecologic and individual-level studies must be the same. In the case where ecologic study is unbiased for ecologic effects and yet still be biased for individual-level effects, then we have the cross-level bias. Originally defined, a cross-level bias is absent when, and only when, β_2 in

the structural equation $Y = a + \beta_1 X_1 + \beta_2 X_2 + e$ (Firebaugh, 1978). In public health literature, Morgenstern (1982) modified the definition to sum of aggregation and specification biases. Greenland and Morgenstern (1989) used cross-level bias as a synonym for ecological bias. Richardson, et. al. (1987) used the term to describe uncontrolled between-group confounding. The said bias will not actually affect the validity of an ecologic test of the null hypothesis, although it still must be considered in interpreting a significant result.

When an aggregated risk model is formed from an individual-level risk model, the two mathematical forms will generally differ, and ignoring this can lead to what has been termed pure specification bias (Greenland, 1992).

The ecological model

$$\mu_k(E) = \exp(\alpha_e + x_k^T \beta_e)$$

is mis-specified whenever the disease model is not linear and $x_k \neq x_{ki}$ for all i, k . The specification bias for β in the ecological model can be quantified with estimating equations. The score equations for the ecological model will be biased because a mis-specified model is being fit.

The above biases should be examined in every application of this design. Guidelines or suggestions on how to overcome the above-mentioned biases can be further formulated.

APPENDIX

Definition 1. The **quadratic exponential model** is given by

$$p_{rk}(y_k; \mu_k, \sigma_k) = \Delta_k^{-1} \exp \{y_k^t \theta_k + w_k^t \lambda_k + c_k(y_k)\} \quad (1)$$

where

$y_k^t = (y_{k1}, \dots, y_{knk})$, $k = 1, \dots, K$ is a sample of K independent multivariate binary observations

$$w_k^t = (y_{k1}y_{k2}, y_{k1}y_{k3}, \dots, y_{k2}y_{k3}, \dots),$$

$$\theta_k^t = \theta_k^t(\mu_k, \sigma_k) = (\theta_{k1}, \dots, \theta_{knk}), \text{ and}$$

$$\lambda_k^t = \lambda_k^t(\mu_k, \sigma_k) = (\lambda_{k11}, \lambda_{k12}, \dots, \lambda_{k22}, \lambda_{k23}, \dots) \text{ are parameters}$$

$\Delta_k = \Delta_k \{\theta_k, \lambda_k, c_k(\cdot)\}$ is a normalization constant defined by

$$\Delta_k = \sum \exp \{y_k^t \theta_k + w_k^t \lambda_k + c_k(y_k)\} \text{ with summation}$$

over all 2^{nk} possible values of y_k

$c_k(\cdot)$ is a "shape" function that can be expressed as a linear combination of products of three or more of the elements of y_k .

Definition 2. Let $A=\{(x, y): f_{x, y}(x, y) > 0\}$ and $B=\{(u, v): u = g_1(x, y) \text{ and } v = g_2(x, y) \text{ for some } (x, y) \in A\}$. Suppose A defines a one-to-one transformation onto B . These x and y can be expressed in terms of u and v so that we have the inverse transformation $x = h_1(u, v)$ and $y = h_2(u, v)$. The determinant of J given by

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

is called the Jacobian of the transformation.

Theorem 1. The score estimating equation for β and α of the quadratic exponential model in equation (1) can be written as

$$K^{-1/2} \sum_{k=1}^K D_k^t V_k^{-1} \mathbf{f}_k = \mathbf{0}. \quad (2)$$

The corresponding information matrix is given by

$$W = K^{-1} \sum_{k=1}^K D_k^t V_k^{-1} D_k \quad (3)$$

where

$$D_k = \begin{pmatrix} \frac{\partial \mu_k}{\partial \beta^t} & 0 \\ \frac{\partial \sigma_k}{\partial \beta^t} & \frac{\partial \sigma_k}{\partial \alpha^t} \end{pmatrix} \quad \mathbf{f}_k = \begin{pmatrix} y_k - \mu_k \\ s_k - \sigma_k \end{pmatrix}$$

$$V_k = \begin{pmatrix} V_{k11} & V_{k12} \\ V_{k21} & V_{k22} \end{pmatrix} = \begin{pmatrix} \text{var } y_k & \text{cov}(y_k, s_k) \\ \text{cov}(s_k, y_k) & \text{var } s_k \end{pmatrix}$$

and where $s_k^t = (s_{k12}, s_{k13}, \dots, s_{k23}, \dots)$, $s_{kij} = (y_{ki} - \mu_{ki})(y_{kj} - \mu_{kj})$ is the vector of covariances

Proof .

Consider the quadratic exponential model in equation (1). For ease of notation, let us drop the subscript k so that $\mathbf{y}^t = (y_1, \dots, y_n)$ is distributed according to

$$p_r(\mathbf{y}) = \Delta^{-1} \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \quad (4)$$

Note that Δ^{-1} is a normalization constant so that the marginal mean $\boldsymbol{\mu}$ is given by

$$\frac{\partial p_r(\mathbf{y})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mu} = \sum_{\mathbf{y}} \mathbf{y} \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1}$$

The product moment $\boldsymbol{\eta} = (\eta_{11}, \eta_{12}, \dots, \eta_{22}, \dots)$, where $\eta_{ij} = \sigma_{ij} + \mu_i \mu_j$ is given by

$$\frac{\partial p_r(\mathbf{y})}{\partial \boldsymbol{\theta}} = \boldsymbol{\eta} = \sum_{\mathbf{y}} \mathbf{w} \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1}$$

Now,

$$\frac{\partial^2 p_r(\mathbf{y})}{\partial \boldsymbol{\theta}^2} = \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} = \sum_{\mathbf{y}} \mathbf{y}^2 \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1} = \text{var } \mathbf{y}$$

$$\frac{\partial^2 p_r(\mathbf{y})}{\partial \boldsymbol{\lambda}^2} = \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} = \sum_{\mathbf{y}} \mathbf{w} \mathbf{y} \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1} = \text{cov}(\mathbf{w}, \mathbf{y})$$

$$\frac{\partial^2 p_r(\mathbf{y})}{\partial \boldsymbol{\theta}^2} = \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}} = \sum_{\mathbf{y}} \mathbf{y} \mathbf{w} \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1} = \text{cov}(\mathbf{y}, \mathbf{w})$$

$$\frac{\partial^2 p_r(\mathbf{y})}{\partial \boldsymbol{\lambda}^2} = \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} = \sum_{\mathbf{y}} \mathbf{w}^2 \exp \{ \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) \} \Delta^{-1} = \text{var } \mathbf{w}$$

The inverse Jacobian of the transformation from canonical parameters $\boldsymbol{\theta}, \boldsymbol{\lambda}$ to marginal parameters $\boldsymbol{\mu}, \boldsymbol{\eta}$ is

$$\tilde{V} = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} & \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}} \\ \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}} & \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \text{var } \mathbf{y} & \text{cov}(\mathbf{y}, \mathbf{w}) \\ \text{cov}(\mathbf{w}, \mathbf{y}) & \text{var } \mathbf{w} \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

Since V is simply the variance matrix for $(\mathbf{y}^t, \mathbf{w}^t)$, the transformation will be one-to-one unless the distribution of $(\mathbf{y}^t, \mathbf{w}^t)$ is degenerate. Using equation (4) to get the log-likelihood, then we have

$$\begin{aligned} p_r(\mathbf{y}) &= \Delta^{-1} \exp\{\mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y})\} \\ \rightarrow \Delta p_r(\mathbf{y}) &= \exp\{\mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y})\} \\ \rightarrow \log \Delta p_r(\mathbf{y}) &= \log \exp\{\mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y})\} \\ \rightarrow \log \Delta + \log p_r(\mathbf{y}) &= \{\mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y})\} \\ \rightarrow \log p_r(\mathbf{y}) &= \{\mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y})\} - \log \Delta \end{aligned}$$

If we let $\ell = \log p_r(\mathbf{y})$, then we have

$$\ell = \mathbf{y}^t \boldsymbol{\theta} + \mathbf{w}^t \boldsymbol{\lambda} + c(\mathbf{y}) - \log \Delta$$

Suppose $\boldsymbol{\beta}$ is the parameter vector in the mean response

$$\boldsymbol{\mu}_k' = \boldsymbol{\mu}_k'(\boldsymbol{\beta}) = \{E(y_{k1}), E(y_{k2}), \dots\}.$$

Also, suppose $\boldsymbol{\alpha}$ is a parameter vector that characterize the variance

$$V_{k11} = V_{k11}(\boldsymbol{\beta}, \boldsymbol{\alpha}).$$

Let

$$\tilde{D}_k = \begin{pmatrix} \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\beta}} & 0 \\ \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\alpha}} \end{pmatrix} \quad \tilde{f}_k = \begin{pmatrix} \mathbf{y} - \boldsymbol{\mu} \\ \mathbf{w} - \boldsymbol{\eta} \end{pmatrix}$$

Then, $\tilde{D}^t \tilde{V}^{-1} \tilde{f} =$

$$\begin{pmatrix} \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\beta}} \\ 0 & \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\alpha}} \end{pmatrix} \begin{pmatrix} \frac{\frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}}}{\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} - \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}}} & \frac{-\frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}}}{\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} - \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}}} \\ \frac{-\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}}}{\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} - \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}}} & \frac{\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}}}{\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\lambda}} - \frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\eta}^t}{\partial \boldsymbol{\theta}}} \end{pmatrix} \begin{pmatrix} \mathbf{y} - \boldsymbol{\mu} \\ \mathbf{w} - \boldsymbol{\eta} \end{pmatrix}$$

$$= \left(\begin{array}{l} \frac{\frac{\partial \mu^t}{\partial \beta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \eta^t}{\partial \beta} \frac{\partial \mu^t}{\partial \lambda}}{\quad} (\mathbf{y} - \boldsymbol{\mu}) + \frac{-\frac{\partial \mu^t}{\partial \beta} \frac{\partial \eta^t}{\partial \theta} + \frac{\partial \eta^t}{\partial \beta} \frac{\partial \mu^t}{\partial \theta}}{\quad} (\mathbf{w} - \boldsymbol{\eta}) \\ \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} \quad \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} \\ -\frac{\frac{\partial \eta}{\partial \alpha} \frac{\partial \mu^t}{\partial \lambda}}{\quad} \quad -\frac{\frac{\partial \eta}{\partial \alpha} \frac{\partial \mu^t}{\partial \theta}}{\quad} \\ \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} (\mathbf{y} - \boldsymbol{\mu}) + \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} (\mathbf{w} - \boldsymbol{\eta}) \\ \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} \quad \frac{\frac{\partial \mu^t}{\partial \theta} \frac{\partial \eta^t}{\partial \lambda} - \frac{\partial \mu^t}{\partial \lambda} \frac{\partial \eta^t}{\partial \theta}}{\quad} \end{array} \right)$$

$$= \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \alpha} \end{pmatrix}$$

by chain rule in partial differentiation

$$= \tilde{\mathbf{D}}^t \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{f}}$$

According to Prentice and Zhao (1991), we can write $\tilde{\mathbf{f}} = \mathbf{U}\mathbf{f}$, $\tilde{\mathbf{V}} = \mathbf{U}\mathbf{V}\mathbf{U}^t$ and $\tilde{\mathbf{D}} = \mathbf{U}\mathbf{D}$, where \mathbf{f} , \mathbf{V} , and \mathbf{D} are, aside from the subscript k , as defined above and the block-diagonal matrix \mathbf{U} is given by

$$\mathbf{U} = \begin{pmatrix} \mathbf{I}_{nk} & \mathbf{0} \\ \mathbf{C}_k & \mathbf{I}_{\{1/2nk(nk-1)\}} \end{pmatrix}$$

where \mathbf{I}_m denotes an identity matrix of dimension m and

$$\mathbf{C}_k^t = (\mathbf{C}_{k1}^t \mid \dots \mid \mathbf{C}_{knk}^t), \text{ where } \mathbf{C}_{ki} = (\mathbf{0}_{ki} \mid \mathbf{X}_{ki} \mid \mu_{ki} \mathbf{I}_{nk-1}),$$

where $\mathbf{0}$ denotes a zero matrix of dimension $(nk-1) \times (i-1)$, which is absent for $i=1$, and $\mathbf{X}_{ki}^t = (\mu_{k,i+1}, \dots, \mu_{knk})$.

Hence, by substitution the score and information can be written as $\tilde{\mathbf{D}}^t \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{f}}$ and $\tilde{\mathbf{D}}^t \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{D}}$, respectively.

REFERENCES

- Anderson, A.B. and Prentice R.L.(1998) On the Accommodation of Disease Rate Correlations in Aggregate Data Studies of Disease Risk Factors. *Biometrics*, **54**, 1527-1540
- Carrol, R.J. (1997) Surprising Effects of Measurement Errors on an Aggregate Data Estimators. *Biometrika*, **84**, 231-234
- Cleave, N., et al. (1995) Evaluation of Methods for Ecological Inference. *Journal of Royal Statistical Society, A*, **158**, 55-72
- Greenberg, et al. (2001) *Medical Epidemiology*, 3rd edition, McGraw-Hills Co., Inc.
- Greenland S. and Robins J. (1994) Invited Commentary: Ecologic Studies – Biases, Misconceptions, and Counterexamples. *American Journal of Epidemiology*, **139**, 747-760
- Guthrie K.A.and Sheppard L. (2001) Overcoming Biases and Misconceptions in Ecological Studies. *Journal of Royal Statistics [A]*, **164**, 141-154
- Guthrie K.A, Sheppard L. and Wakefield J. (2002) A Hierarchical Aggregate Data Model with Spatially Correlated Disease Rates. *Biometrics*, **58**, 898-905
- Hammond J.L.(1973) Two Sources of Error in Ecological Correlations. *American Sociological Review*, **38**, 764-777
- King, G. (1997) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, Princeton University Press, Princeton, New Jersey.
- Lawson A. B. and Denison D. G. T. (2002) *Spatial Cluster Modeling*, Chapman and Hall/ CRC, Florida USA
- Liang K. and Zeger S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22
- Lindsey, J.K. (1993) *Models for Repeated Measurement*, Oxford University Press, Inc., New York USA
- Plummer, M. and Clayton, D.(1996) Estimation of Population Exposure in Ecological Studies. *Journal of Royal Statistical Society, B*, **58**, 113-126
- Prentice R.L. and Sheppard, L. (1995) Aggregate Data Studies of Disease Risk Factors. *Biometrika*, **82**, 113-125
- Prentice R.L. and Zhao L.P. (1991) Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. *Biometrics*, **47**, 825-839
- Sheppard L. (2003) Insights on Bias and Information in Group-Level Studies. *Biometrics*, **4**, 265-278
- Sheppard L. and Prentice R.L. (1995) On the Realibility and Precision of Within-and Between-Population Estimates of Relative Rate Parameters. *Biometrics*, **51**, 853-863
- Sheppard L., Prentice R.L. and Rossing M. (1996) Design Considerations for Estimation of Exposure Effects on Disease Risk Using Aggregate Data Studies. *Statistics of Medicine*, **15**, 1849-1858
- Steel D.G, Tranmer, M. and Holt D. *Logistic Regression Analysis with Aggregate Data: Tackling the Ecological Fallacy* (unpublished yet).
- Wakefield J. (2003) *A Critique of Statistical Aspects of Ecological Studies in Spatial Epidemiology* (to appear in *Ecological and Environmental Statistics*)
- Wakefield, J. and Salway, R. (2001) A Statistical Framework for Ecological and Aggregate Data Studies. *Journal of Royal Statistical Society*, **164**, 119-137
- Zhao L.P. and Prentice R.L. (1990) Correlated Binary Regression Using a Quadratic Exponential Model. *Biometrika*, **77**, 642-648