WITHIN CLUSTER RESAMPLING

by

Elaine Borland Hoffman

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2191T

May 1998

WITHIN CLUSTER RESAMPLING

by
Elaine Borland Hoffman

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, School of Public Health.

Chapel Hill
1998

Approved by

Chair: Lawrence L. Kupper

Advisor: Clarice R. Weinberg

Reader: Pranab K. Sen

Reader: Irva Hertz-Picciotto

Reader: Marie Davidian

ABSTRACT

ELAINE BORLAND HOFFMAN: Within Cluster Resampling
(Under the direction of Clarice R. Weinberg)

Dependence among observations from the same group is known as within-cluster correlation. Typically, distinct clusters are independent, but special methods are required to account for the existence of the within-cluster correlation. Although not accounting for the within-cluster correlation may produce consistent parameter estimates, the variances will be under-estimated. Within Cluster Resampling (WCR) is proposed as a method for analyzing any clustered data; however, this dissertation will focus on clustered binary data.

The proposed idea is to sample one observation from each cluster. A set of independent observations called a resampled data set results, which simply can be analyzed by standard software. As the name Within Cluster Resampling implies, the above process is repeated by randomly sampling with replacement one observation from each cluster. A series of data sets is created by resampling a large number of times. The regression parameter is estimated by the average of the resample-based estimates. The proposed variance will account for the correlation among observations from the same cluster.

The interpretation of a WCR parameter is the population-averaged difference associated with a unit change in a covariate corresponding to a randomly selected observation from a randomly selected cluster. WCR is unlike traditional marginal analysis methods. WCR is a cluster-based method that equally weights each cluster, whereas traditional marginal methods are observation-based and the weight of the cluster increases with its size.

Simulation results are presented that demonstrate the finite sample behavior of the WCR method for analyzing clustered binary data. The simulated data sets had varying amounts of within-cluster correlation as well as equal and unequal cluster sizes. For comparison, each non-informative cluster size data set was analyzed by GEE with an appropriate user-specified, working correlation matrix.

The dissertation includes a thorough literature review of clustered binary data, a proof of the asymptotic normality of a standardized WCR estimator and the consistency of the proposed variance estimator, simulations comparing WCR with GEE, and analyses of actual data.

ACKNOWLEDGMENTS

I want to thank my family and friends for their encouragement throughout my academic career. In particular, I praise my husband, Matt, for his constant love and patience with me throughout this roller coaster ride they call graduate school. Without Matt, my success would not be as great and my life would not be as full. Also a special thanks to my friend Lisa Tomasko, we supported each other on our separate but similar endeavors and accomplished what a long time ago seemed insurmountable.

In addition, I want to express my sincere gratitude to Dr. Clarice Weinberg of NIEHS for advising me as her first biostatistics doctoral student, allowing me to research an incredibly interesting and important statistical topic, and spending so much time over the past 2 years helping me with the research. My dissertation has been a very positive experience and I am a better statistician for having been mentored by Clare.

I would also like to recognize my committee, Dr. P.K. Sen, Dr. Larry Kupper, Dr. Irva Hertz-Picciotto, and Dr. Marie Davidian for their time and guidance throughout the dissertation process.

I am truly grateful to NIEHS for their financial assistance throughout my graduate career. First, I was a trainee on the NIEHS Environmental Training Grant #ES07018 through Dr. Larry Kupper in the Department of Biostatistics at UNC. Secondly, my dissertation research was supported in full by a pre-doctoral IRTA which I received through Dr. Clarice Weinberg.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CLT	Central Limit Theorem
CS	Cluster-specific
diag	diagonal
EM	Expectation/Maximization
$ese(\cdot)$	empirical standard error
exch.	exchangeable
exp	exponential
GEE	Generalized Estimating Equations
GLM	Generalized Linear Model
ind.	independent
IRLS	Iteratively Reweighted Least Squares
IWLS	Iteratively Weighted Least Squares
ln	natural logarithm
MAR	Missing at Random
MCAR	Missing Completely at Random
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MQLE	Maximum Quasi-likelihood Estimation
MVN	Multivariate Normal
OLR	Ordinary Logistic Regression
OR	Odds-ratio
OS	Observation-specific
PL	Pseudo-likelihood
PQL	Penalized Quasi-likelihood

$std(\cdot)$	standard deviation
WCR	Within Cluster Resampling

CHAPTER I

INTRODUCTION

Often, there is dependence among the observations from the same group, a phenomenon known as within-cluster correlation. Although distinct clusters can often be assumed to be independent, because of the dependency within clusters, methods for analyzing independent data are not suitable for clustered data. For the purposes of this dissertation, I will focus on clustered binary data. However, the method to be proposed is general and can be applied to any clustered data.

The leading methods for analyzing clustered binary data include marginal (population-averaged) models, random-effects (subject-specific) models, conditional models, and response-conditional models. However, these methods have subtle differences in the underlying model, the required assumptions, and the interpretation of the parameters. The choice of one method over another depends upon the sampling scenario, the question of interest, and the plausibility of the required assumptions.

1.1 Terminology and Notation

To be clear and consistent throughout, the group of dependent binary outcomes will be referred to as a "cluster" and indexed by i. A cluster in an application may be specific to an individual, an animal, a plot of land, etc. The binary outcome nested within i will be referred to as an "observation" and indexed by j. Some examples of outcomes are whether or not a particular pregnancy of a woman resulted in a live birth, whether or

not a pup from a litter has a birth-defect, or whether or not a plant from a plot of land has a certain disease. Covariates that pertain to the cluster and do not vary among observations within the cluster will be called cluster-specific covariates, and covariates that pertain to an observation will be called observation-specific covariates.

A cluster may be a set of outcomes recorded over time, where the outcomes within a cluster have a natural ordering. A cluster may also have structured, but not time-ordered, outcomes, such as cataract or no cataract in the left and right eyes, or ear infections in none, one or both ears. Another cluster may be the set of binary outcomes for a litter of pups, where the pup outcomes within a litter are random and unordered (unless birth order is important).

The random vector of binary outcomes for cluster i will be denoted by $\underset{\sim}{Y}_i$, the value or realization of the binary variable vector will be represented by $\underset{\sim}{y}_i$, and $\underset{\sim}{X}_i$ will be the covariate matrix associated with cluster i, where $i = 1, 2, \ldots, I$. In general, one dimension of $\underset{\sim}{X}_i$, the number of covariates, will be fixed by design while the other, corresponding to the number of observations within the cluster, may be random or partially random due to a structured number of observations that is then subject to missing data. Let Y_{ij}, where $j = 1, 2, \ldots, n_i$, represent observation j of cluster i. In the simplest case, a cluster has a single binary outcome, $n_i = 1$. The most common way to handle the case of $n_i = 1$ is by ordinary logistic regression (OLR) (Cox, 1970). The next case is where each cluster has two observations, $n_i = 2$ for all i. As mentioned earlier, a study might look at cataracts in eyes or ear infections of individuals who might vary in their susceptibility for the outcome of interest. It should be noted that all cluster sizes can be equal, but this is not a necessary condition for the method to be proposed.

1.2 Proposed Method

The method proposed in this paper is Within Cluster Resampling (WCR), a semi-parametric, cluster-based, marginal analysis method. The WCR method can be applied to any set of clustered binary data. The dependency among the outcomes within a cluster can arise simply because the baseline risk varies across clusters or because of heterogeneity across clusters in response to an exposure. Also, it can arise directly, as when infection in one eye spreads to the other. The method to be discussed can apply under any of these paradigms for dependency.

The proposed idea is to analyze clustered binary data by randomly sampling one observation, $y_{i\pi(n_i)}$, from each cluster outcome vector, $\underset{\sim}{y}_i$, where $i = 1, \ldots, I$ denotes the clusters and $\pi(n_i)$is random on the integers $1, \ldots, n_i$. This set of independent observations (one observation from each cluster) can be analyzed easily using standard software; the analysis will produce consistent estimates, $\underset{\sim}{\widehat{\beta}}$, of the effects of interest, $\underset{\sim}{\beta}$. As the name of the proposed method indicates, Within Cluster Resampling, the above procedure is repeated by randomly sampling with replacement from each cluster. By resampling a finite number of times, say Q, a series of Q data sets of size I is created. Each of the "resampled data sets" consists of independent observations and can be analyzed by a generalized linear model (GLM) (such as the logistic model) using standard software. For example, with pregnancy history data, each resampled data set includes one pregnancy from each couple together with the associated covariates for that pregnancy. Each resampled data set will be denoted by $R(q)$, where $q = 1, \ldots, Q$ denotes the number of resamples, and the vector of consistent estimates from each resample will be denoted by $\underset{\sim}{\widehat{\beta}}(R; q)$. The overall regression parameter vector is estimated by the average of the vectors of resample-based estimates, $\underset{\sim}{\overline{\beta}} = \sum_{q=1}^{Q} \frac{\underset{\sim}{\widehat{\beta}}(R;q)}{Q}$. The variance to be proposed for $\underset{\sim}{\overline{\beta}}$ will account for the correlation among observations from the same cluster. The WCR method is schematically displayed in Figure 1.1.

Figure 1.1. Resampling Scheme for WCR

Within Cluster Resampling Parameter Estimation:

Consider a hypothetical data set consisting of I clusters ...

$$\underset{\sim}{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \end{bmatrix} \quad \underset{\sim}{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} \quad \underset{\sim}{y}_3 = [y_{31}] \quad \underset{\sim}{y}_4 = \begin{bmatrix} y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{bmatrix} \quad \cdots \quad \underset{\sim}{y}_I = \begin{bmatrix} y_{I1} \\ y_{I2} \end{bmatrix}$$

By repeated (Q times), random sampling of one observation from each of the above I clusters, a series of Qresampled data sets $R(q)$, $q = 1, \ldots, Q$, of size I and resample-based estimates ($\widehat{\underset{\sim}{\beta}}\ (R; q)$) are obtained...

$$R(1) = \begin{vmatrix} y_{11}, \underset{\sim}{x}_{11} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{44}, \underset{\sim}{x}_{44} \\ \vdots \\ y_{I1}, \underset{\sim}{x}_{I1} \end{vmatrix} \xrightarrow{\text{GLM}} \widehat{\underset{\sim}{\beta}}\ (R; 1) \qquad \cdots \qquad R(Q) = \begin{vmatrix} y_{12}, \underset{\sim}{x}_{12} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{42}, \underset{\sim}{x}_{42} \\ \vdots \\ y_{I2}, \underset{\sim}{x}_{I2} \end{vmatrix} \xrightarrow{\text{GLM}} \widehat{\underset{\sim}{\beta}}\ (R; Q)$$

The resample-based estimates are combined into an overall estimate ($\overline{\underset{\sim}{\beta}}$) by averaging across the resamples...

$$\overline{\underset{\sim}{\beta}} = \frac{\sum_{q=1}^{Q} \widehat{\underset{\sim}{\beta}}(R;q)}{Q}$$

Conditional on the observed data, the Q estimators, $\widehat{\underset{\sim}{\beta}}\ (R; q)$, $q = 1, \ldots, Q$, are independent and identically distributed. Their distribution will be referred to as the resampling distribution; and, by the Central Limit Theorem (CLT), $\overline{\underset{\sim}{\beta}} \xrightarrow{p} \underset{\sim}{\beta}$ consistently (Sen and Singer, 1993). To approximate adequately the first two moments of the resampling distribution, Q may need to be large.

A litter-effects example will attempt to elucidate the difference between the WCR parameter and the traditional marginal parameter. One difference is that WCR is a *cluster-based*, population-averaged approach, whereas, a traditional marginal method, such as generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986), is an *observation-based*, population-averaged approach. Therefore, in litter-effects problems, an observation-based approach is influenced by the fertility of a dam, but a cluster-based approach is not. This means that litters from fertile dams are given greater weight in an observation-based analysis, but the WCR method gives equal weight to each litter (cluster) in an analysis, regardless of the fertility of a dam. The WCR approach can be thought of as an "unweighted" analysis approach, whereas, the traditional marginal approach is a "weighted" analysis approach. For illustrative purposes, consider a teratological experiment at the Shell Toxicology Laboratory, Sittingbourne (Paul, 1982). This litter-effects study was interested in the effect of a compound on the live fetuses of 84 dams. The compound had four dose levels: control, low, medium, and high. The outcome of interest is the number of abnormal, live fetuses from each litter. The data for the low dose group are as follows:

abnormal, live fetuses	0	1	1	0	2	0	1	0	1	0	0	3	0	0	1	5	0	0	3
total live fetuses	5	11	7	9	12	8	6	7	6	4	6	9	6	7	5	9	1	6	9

The traditional marginal risk from an observation-based approach is the probability of an abnormal random fetus, which is estimated by the total number of events divided by the total number of observations:

$$Pr_{(\mathrm{TM})}(\text{abnormal}) = \frac{\text{total number of abnormal fetuses}}{\text{total number of fetuses}}$$

$$= \frac{(0+1+1+0+2+0+1+0+1+0+0+3+0+0+1+5+0+0+3)}{(5+11+7+9+12+8+6+7+6+4+6+9+6+7+5+9+1+6+9)}$$

$$= \frac{18}{133} = 0.135.$$

The marginal risk from WCR, a cluster-based approach, is the probability of an abnormal random fetus from a random litter, which is estimated by averaging the litter-specific probabilities of abnormalities:

$$Pr_{(\mathrm{WCR})}(\text{abnormal}) = \frac{\text{sum of litter-specific abnormality probabilities}}{\text{number of litters}}$$

$$= \frac{(\frac{0}{5}+\frac{1}{11}+\frac{1}{7}+\frac{0}{9}+\frac{2}{12}+\frac{0}{8}+\frac{1}{6}+\frac{0}{7}+\frac{1}{6}+\frac{0}{4}+\frac{0}{6}+\frac{3}{9}+\frac{0}{6}+\frac{0}{7}+\frac{1}{5}+\frac{5}{9}+\frac{0}{1}+\frac{0}{6}+\frac{3}{6})}{19}$$

$$= 0.113.$$

Because WCR equally weights each litter, the WCR risk estimate is smaller than the traditional marginal risk estimate since the smaller litters in this example have fewer abnormal, live fetuses. The traditional marginal approach weights the large litters more.

From this example, it should be clear that WCR is cluster-based and the traditional marginal approach is observation-based. Also, WCR is not estimating the same parameter as an observation-based, population-averaged analysis; at times, the two parameters are quite different. In the case of equal sized clusters, the two approaches are estimating the same parameter, but in very different ways. Therefore, a comparison between WCR analyses and observation-based analyses performed on the same data is difficult, as well as somewhat illogical. Choosing between a WCR analysis as opposed to an observation-based, marginal analysis depends upon the question of interest, because the two analyses are attempting to answer different questions.

An example using pregnancy history data will help to clarify the interpretation of a parameter from the WCR method. Suppose a study investigates the relationship between cigarette smoking and spontaneous abortion in women of child-bearing age. The data for each particular woman consist of whether she had smoked or not during each of her pregnancies, and which pregnancies, if any, resulted in a spontaneous abortion. The resampling, as described above, is carried out and a logistic model is fit to each resampled

data set. The interpretation of each resample-based parameter estimate for smoking, $\widehat{\beta}(R;q)_{smoke}$, $q = 1, \ldots, Q$, is as the estimate of the natural logarithm of a marginal odds ratio. The interpretation of the WCR parameter estimate, $\overline{\beta}_{smoke}$, is the difference in log odds for spontaneous abortion between a randomly sampled smoking pregnancy from a randomly selected woman and a randomly sampled non-smoking pregnancy from a randomly selected woman. This interpretation of the WCR parameter estimate reflects the one-per-cluster sampling scheme of the procedure.

The interpretation of a parameter from WCR differs from that of more traditional parameters. Like random-effects models, WCR is a cluster-based analysis method, but without an assumed distribution for the heterogeneity among clusters. If the premise is that each cluster has its own inherent risk, the mean risk from WCR, μ_{WCR}, is the average of these risks across the population of *clusters*. In contrast to the mean risk from WCR, the traditional marginal mean risk, μ_{TM}, is the average risk across the population of *observations* not clusters. WCR has an additional level of sampling in that, with WCR, the unit of analysis is the cluster, from which a random observation is selected as a representative. This contrasts with the traditional marginal analysis, e.g. GEE, where the unit of analysis is the observation.

The potential advantages of WCR are the intuitive interpretation of the parameters and the easy implementation of WCR with existing software. In addition, a WCR analysis equally weights each cluster. This may lead to valid inference if the cluster size is "informative". Informative cluster sizes occur when the expected number of observations within a cluster is correlated with the baseline risk. An example is pregnancy history data, where women at higher risk of miscarriage may have more pregnancies than women who reproduce successfully (Gladen, 1986).

This advantage of WCR over a traditional marginal model, for certain contexts, is illustrated by a litter-effects scenario. Research is often performed where the species of animals or the age-groups of people vary by study. After several studies have been

performed, researchers may want to compare the results across studies (species). If there are informative cluster sizes, then traditional marginal analyses may not be valid because they tend to implicitly assume there is no relationship between the cluster size and the marginal risk. Therefore, comparisons between the studies (species) may be inappropriate. Consider a litter-effects example. Suppose there are two species of dams (A and B), and each includes a 50/50 mix of two different baseline risks for a pup outcome of interest. The probability of a pup having the outcome in the lower risk group (r) is denoted by p_r and in the higher risk group (r^*) is denoted by p_{r^*} for both species. Species A has litters of size two, regardless of risk group, whereas species B has litters of size two in the lower risk group and litters of size one in the higher risk group. [This is an unrealistic, purely heuristic example in that the cluster size is highly informative: the size of the cluster is perfectly correlated with the baseline risk for species B.] The species-based mean risk based on WCR for both species A and B is the average risk across the population of litters, $\mu_{WCR(\mathrm{A})} = \mu_{WCR(\mathrm{B})} = \frac{p_r + p_{r^*}}{2}$. By contrast, the traditional marginal models, because they are observation-based and not cluster-based, do not give a consistent estimate of the species-based marginal mean, except in the trivial case where there is no heterogeneity of risk ($p_r = p_{r^*}$). The marginal mean risk for the traditional marginal model for species A is $\mu_{PA(\mathrm{A})} = \frac{p_r + p_{r^*}}{2}$ and for species B is $\mu_{PA(\mathrm{B})} = \frac{2p_r + p_{r^*}}{3}$. Rather than estimating the species-based mean risk, the traditional marginal approaches give greater weight to the litters of size two than to the litters of size one in species B. As a result, the traditional marginal approaches are not consistent for the species-based mean risk, although the species-based mean risks are consistent using WCR. This is an important point since informative cluster sizes are often not known when analyzing a data set as well as for comparing populations with different degrees of informative cluster sizes.

1.3 Dissertation Proposal

The proposed research will include a thorough review of the clustered binary data literature, a proof of the asymptotic normality of a standardized version of the WCR estimator, the consistency of the proposed variance estimator, analyses of real data sets by the WCR method, and simulations comparing WCR with GEE.

The literature review will verify that the WCR method is new, and will elucidate weaknesses in the current methods that may be addressed by the proposed method.

The asymptotic normality of a standardized version of the WCR estimator will then be proven. The approach contemplated is as follows. We approximate each resample-based estimator in terms of a function of the y_{ij}'s. Since the WCR estimator is the average of the resample-based estimators, it will be possible to group the outcomes from the same cluster together. Therefore, the standardized version of the WCR estimator can be rewritten in terms of a function of independent observations so that a Central Limit Theorem will apply.

The WCR method offers advantages over the current methods. One robustness advantage over random-effects models is that it does not assume a parametric representation for the heterogeneity in the baseline risk. Also, unlike GEE methods (Liang and Zeger, 1986; Zeger and Liang, 1986), the WCR method accounts for the within-cluster correlation without the need to specify a working correlation matrix.

Simulations have been planned to explore the finite sample behavior of the WCR method. Also, these simulations will help assess the validity of the proposed variance formula and asymptotic normality under a variety of situations. The simulations will be used to characterize the statistical properties of the method as Q increases, in order to rule out the possibility that the conjectured variance, discussed in Chapter III, is incorrect.

To compare the performance of the WCR method with other standard methods, the simulated data sets will be analyzed by WCR and by GEE for non-informative cluster sizes. The results will potentially provide further evidence that WCR is a valid and widely

applicable method for analyzing clustered binary data. The simulations will encompass equal, unequal, and informative cluster sizes for several sample sizes. Also, the simulations will include cluster-specific and observation-specific covariates, as well as various amounts of within-cluster correlation.

There are many overlooked issues relating to the source of the dependency among a cluster's outcomes. Dependency can arise as a result of the presence of heterogeneity in the baseline, which is the traditional assumption. It also may be a feature of the study design, such as auto-correlation in longitudinal studies; it may arise from the effect of a contagion, such as the right eye contracting a "disease" from the left eye; or it may result from heterogeneous susceptibility, which means that every cluster or unit responds differently to an exposure. How these different sources of dependency affect an analysis and the interpretation of the parameters may be examined in the future.

Other future work may include a paired resampling method for subject-specific parameter estimation. This approach would be similar to conditional logistic regression. However, the paired resampling would potentially allow subject-specific parameters to be estimated without specialized and CPU-intensive software. The complexity and feasibility of paired resampling has not been discussed extensively. Another area of future research may involve the often-ignored issue relating to the effects of informative cluster sizes.

Therefore, the proposal for this dissertation is to thoroughly review the relevant clustered binary data literature, prove a theorem about the asymptotic normality of a standardized version of the WCR estimator, analyze some real data examples, and compare WCR with the existing methods via simulations. In the future, I may extend this work by exploring subject-specific parameter estimation via paired resampling, as well as the effects of different sources of dependency and informative cluster sizes.

CHAPTER II

LITERATURE REVIEW

2.1 Background

2.1.1 Ordinary Logistic Regression

In the simplest case, where $n_i = 1$ for all i, the most common method for the analysis of binary data is logistic regression. Logistic regression uses the logit link because it is the canonical link for Bernoulli data (McCullagh and Nelder, 1989). Other links such as the identity, complementary log-log, or probit are also possible choices for binary data. The interpretation of the parameters from models with the logit link is as log odds-ratios associated with a unit change in the covariate.

If Y_{ij} is a binary outcome variable and x_i is a scalar covariate for cluster i, $i = 1, \ldots, I$, then the probability of a success can be denoted as $\pi(x_i) = E(Y_{ij}|x_i) = Pr(y_{ij} = 1|x_i)$. If clusters are all size one, we have independent binary variables $Y_1, \ldots, Y_I$, and the likelihood function is $l(\beta) = \prod_{i=1}^{I} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}$. More generally, the n_i will not all be one and the contribution to the likelihood from each cluster is a probability from the binomial distribution. In logistic regression, the covariates are related to the outcome variable by $\pi(x_i) = \frac{exp(\beta_0+\beta_1 x_i)}{1+exp(\beta_0+\beta_1 x_i)}$. By taking the natural logarithm, the logit can be shown to be linear in the parameters, $ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_i$. This is a desirable property since the left-hand side is a smooth, monotonic transformation of the mean of the response. The likelihood for binary

data is maximized by iteratively solving non-linear equations that require updating the weights at each iteration to obtain the parameter estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$ (Fisher, 1935; Bliss, 1935; and Finney, 1971).

2.1.2 Extension to Multiple Outcome Methods

If there are multiple observations for a cluster, such as observations made on pups within a litter, the correlation between the observations (pups) within a cluster (litter) is likely to be greater than zero (Williams, 1975; Gladen, 1979; Haseman and Kupper, 1979; Kupper, Portier, Hogan, and Yamamoto, 1986; Chen and Kodell, 1989). In order to exploit the data fully and to obtain consistent variance estimates, the within-cluster correlation should be taken into account. However, the most naive approach is to ignore it. Ignoring within-cluster correlation provides consistent estimates of the regression parameters if the correlation structure is exchangeable (equal correlation between all pairs of observations within a cluster) and secondary to heterogeneity of risk (Zeger and Liang, 1992). However, the standard errors are under-estimated and the Type I error rate is elevated. The incorrect variance is due to extra-binomial variation, which means that there is more variability among clusters than would be expected based on the binomial (Haseman and Kupper, 1979). In most situations, clustered binary data analysis approaches, such as the WCR method, that account for the within-cluster correlation should be used instead of the naive approach that falsely assumes independence.

2.2 Overview of Clustered Binary Data Models

2.2.1 Marginal Models

Marginal models are a common approach for analyzing clustered binary data when the dependency is not itself of particular interest. However, there are marginal models that explicitly estimate the dependency parameters (Prentice, 1988; Zhao and Prentice, 1990; Prentice and Zhao, 1991; Fitzmaurice and Laird, 1993; Carey, Zeger, and Diggle,

1993). Traditionally, however, marginal models relate the outcome variable to the covariates without explicitly modeling the dependency structure. A general population-averaged model can be written as

$$h(\mu_{ij}) = \underset{\sim}{x}_{ij}^{T}\underset{\sim}{\beta} \qquad \text{and} \qquad var(y_{ij}|\,\underset{\sim}{x}_{ij}) = g(\mu_{ij})\cdot\phi \tag{2.1}$$

where $\mathrm{E}(y_{ij}|\,\underset{\sim}{x}_{ij}) = \mu_{ij}$ is the marginal expectation, h and g are the respective link and variance functions, and ϕ is a scale parameter (Zeger, Liang and Albert, 1988). More will be said regarding marginal models in Section 2.3.

2.2.2 Random-Effects Models

In contrast to the above marginal model, a general random-effects or subject-specific model is given by

$$h^{*}(\nu_{ij}) = \underset{\sim}{x}_{ij}^{T}\underset{\sim}{\beta}^{*} + \underset{\sim}{z}_{ij}^{T}\underset{\sim}{b}_{i} \quad \text{and} \quad var(y_{ij}|\underset{\sim}{b}_{i}, \underset{\sim}{x}_{ij}, \underset{\sim}{z}_{ij}) = g^{*}(\nu_{ij})\cdot\phi \tag{2.2}$$

where $E(y_{ij}|\underset{\sim}{b}_{i}, \underset{\sim}{x}_{ij}, \underset{\sim}{z}_{ij}) = \nu_{ij}, \underset{\sim}{z}_{ij}$ is usually a subset of $\underset{\sim}{x}_{ij}$, and h^{*} and g^{*} are the link and variance functions, respectively (Zeger, Liang and Albert, 1988). The random effect for cluster i, $\underset{\sim}{b}_{i}$, is distributed in general as $F(\underset{\sim}{0}, \underset{\sim}{B})$. The marginal parameters, $\underset{\sim}{\beta}$, from the population-averaged model (2.1), are not the same as the subject-specific parameters, $\underset{\sim}{\beta}^{*}$. Under the random-effects model, the marginal mean specific to covariate vector $\underset{\sim}{x}_{ij}$ is

$$\mu_{ij} = E(y_{ij}|\underset{\sim}{x}_{ij}) = E[E(y_{ij}|\underset{\sim}{b}_i, \underset{\sim}{x}_{ij})] = \int h^{*-1}(\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}^* + \underset{\sim}{z}_{ij}^T\underset{\sim}{b}_i)dF(\underset{\sim}{b}_i|\underset{\sim}{x}_{ij})$$
$$\neq h^{*-1}(\underset{\sim}{x}_{ij}^T\underset{\sim}{\beta}^*), \qquad (2.3)$$

except when h^* is the identity link or $\underset{\sim}{b}_i$ is identically 0.

Another common assumption in subject-specific models is conditional independence. This means that the observations within a cluster are assumed to be independent given the value of the random-effect. This is not always a valid assumption. Random-effects models will be discussed in more detail in Section 2.4.

2.2.3 Conditional Models

Yet another approach for analyzing clustered binary data arises in the context of a study with time-ordering. This is the "conditional" model (Bonney, 1987). A general conditional model can be written as

$$h^{**}(\upsilon_{it}) = \underset{\sim}{x}_{it}^T\underset{\sim}{\beta}^{**} + \underset{\sim}{y}_i^{(t-1)}\underset{\sim}{\alpha}_t \text{ and } var(y_{it}|\, y_{i1}, \ldots, y_{it-1}, \underset{\sim}{x}_{it}) = g^{**}(\upsilon_{it}) \cdot \phi \quad (2.4)$$

where $E(y_{it}|y_{i1}, \ldots, y_{it-1}, \underset{\sim}{x}_{it}) = \upsilon_{it}$, h^{**} and g^{**} are the respective link and variance functions, $\underset{\sim}{y}_i^{(t-1)} = (y_{i1}, \ldots, y_{it-1})$ is a vector of previous outcomes, and $t = 1, \ldots, T$ indexes time. The "response-conditional" model (Rosner, 1984; Qu, et al., 1987; Connolly and Liang, 1990) depends on all of the other outcomes instead of just the past outcomes as in the conditional model proposed by Bonney (1987). Conditional and response-conditional models will be discussed in Section 2.5.

2.2.4 General Comparison of Models

The three "β" parameters from the above models are usually different. The marginal parameter, $\underset{\sim}{\beta}$, in (2.1) is the population-averaged difference (observation-based

in the h scale) corresponding to a unit change in the covariate, whereas, the parameter from the random-effects model, $\underset{\sim}{\beta}^{*}$, in (2.2) represents the within-cluster difference corresponding to a unit change in the covariate. The conditional parameter, $\underset{\sim}{\beta}^{**}$, in (2.4) represents the average difference corresponding to a unit increase in the covariate, given the values of the previous responses (Neuhaus, 1992). Note that this model assumes that $\underset{\sim}{\beta}^{**}$does not depend on the number of previous responses observed. If $T = 1$ then $\underset{\sim}{\beta}^{**}$ has the same population-averaged interpretation as model (2.1).

Each distinct approach is more advantageous in some situations than in others. It may be desirable to use the marginal approach for studies that are interested in population-averaged effects rather than effects within clusters. A random-effects model may be more appropriate when the primary interest is on response within clusters, such as changes over time. Conditional models may be reasonable for data that have a natural ordering, such as longitudinal data, but otherwise the interpretation of the parameters is not clear and the assumption that $\underset{\sim}{\beta}^{**}$ does not depend on t, the number of prior responses, is a strong one.

How the within-cluster correlation is accounted for depends on the choice of model. Although marginal models are affected by the baseline heterogeneity in the population, there is no parametric assumption about this distribution as is required in most random-effects models. The variance, if properly estimated, accounts for the within-cluster correlation in marginal models. In random-effect (subject-specific) models, the within-cluster correlation has a specific structure that arises from the assumed distribution of the random effect. Conditional models account for the within-cluster correlation by conditioning on the history of a cluster and then modeling the effect of that history on the mean response. Since these models depend on the other observations within a cluster, it is not clear that one can assume $\underset{\sim}{\beta}^{**}$ does not depend on time.

Marginal models require fewer assumptions and are easy to implement and interpret. Population-averaged models, unlike random-effects models, cannot however estimate the effect of a given covariate change within a cluster, which may be of interest in longitudinal studies (Neuhaus, Kalbfleisch, and Hauck, 1991) and may be of more importance for personal decision-making (What would quitting smoking do to my personal risk?). A disadvantage of the marginal approach is that separate analyses in two populations with different heterogeneity distributions can produce different parameter values (Zeger, Liang, and Albert, 1988), even if the effects are the same at the cluster-specific level. Also, the population-averaged parameter itself may be much closer to the null than a subject-specific parameter and this can obscure important effects. As already discussed, another potential weakness of the marginal model is that it may be invalid for data with informative cluster sizes.

Subject-specific models require sufficiently large numbers of clusters to estimate the random-effect parameters properly. However, when the number of observations per cluster is too large, the computations become unfeasible. In a random-effects model, the interpretation of a parameter for a cluster-specific covariate is not really subject-specific: the interpretation of "β" in a random-effects model nominally corresponds to the within-cluster effect of a covariate, whereas in the data at hand the covariate does not in fact change its value within any cluster. As discussed above, in such a circumstance the random-effects model parameter must be thought of as having a population-averaged interpretation. Any subject-specific interpretation requires a drastic extrapolation from the data (Galbraith, 1991). If instead some clusters have covariates that change value within the cluster and other clusters do not have changing covariate values, then the interpretation of parameters from a subject-specific model is unclear. There is only a clear interpretation in random-effects models when the covariate is observation-specific and there is some variation in the covariate within each cluster studied. This point is often not appreciated (Zeger, Liang, and Albert, 1988; Galbraith, 1991).

Unlike population-averaged models and like random-effects models, conditional models can take advantage of clustered binary data by modeling within-cluster changes (Ware, Lipsitz, and Speizer, 1988). Criticism of conditional models often relates to the interpretation of parameters. Parameters from conditional models are not marginal or subject-specific, but their interpretation depends on the values of the other outcomes. Conditional models based on time-ordering build on a mathematically convenient specification of the joint probability as the product of successive, conditional probabilities. Each conditional probability is often modeled logistically as a function of the covariates and previous outcomes. Although this is a statistically convenient representation, it does not offer intuitive appeal (Prentice, 1988). Another concern is that conditioning explains away part of the effects of the covariates (Weinberg, 1993). Consequently, the estimated effects of the covariates of interest are lessened when prior outcomes are used as covariates. This partially defeats the point of finding the magnitude of the relationship between covariates and the response, unless the goal of an analysis is purely prognostic: to predict a future outcome based upon previous outcomes.

Conditional parameters tend to be smaller in absolute value than marginal parameters, and marginal parameters tend to be smaller in absolute value than the corresponding random-effect parameters when there is positive within-cluster and covariate correlations (Zeger, Liang and Albert, 1988; Rosner and Milton, 1988; Neuhaus and Jewell, 1990; Neuhaus, Kalbfleisch, and Hauck, 1991; Neuhaus, 1992). The difference between the marginal and the random-effect parameters increases as the heterogeneity among clusters increases, provided there is variation in the exposure within clusters (Neuhaus, 1992). When there is no covariate effect, all of the models estimate the same null effect (Neuhaus, 1992). The test of the null hypothesis should have the same Type I error rate, regardless of whether a marginal, subject-specific, or conditional model is applied.

Thus far, only general advantages and disadvantages of marginal, random-effects, and conditional (response-conditional) models have been discussed. There are many ways to implement each of these general models. Some of the main approaches of marginal, random-effects, conditional, and response-conditional models will be discussed further next.

2.3 Marginal Models

2.3.1 Introduction

Numerous ways have been developed to fit a marginal model. Although one approach may require all covariates to be cluster-specific, another may require an exchangeable correlation structure or limit the cluster size to two. Within the class of marginal models, the research context often determines the optimal approach. Marginal models discussed in this section are categorized into two main groups: maximum likelihood and quasi-likelihood methods. A third group which includes a bootstrap method, is also briefly considered in this section.

2.3.2 Maximum Likelihood Methods

The beta-binomial distribution has been extensively researched, especially for the litter-effects problem, by Williams (1975, 1982), Crowder (1978), Haseman and Kupper (1979), Segreti and Munson (1981), Paul (1982), and Prentice (1986)). If the outcome of interest, Y_{ij}, is a binary outcome variable, then each observation within a cluster can be viewed as a Bernoulli trial with some probability of success specific to the cluster (litter). If outcomes are independent, conditional on a cluster, the sum of the number of successes within a cluster has a binomial distribution. If the probability of a success for an observation within a cluster is assumed to be p_i, where i indexes the cluster, and p_i is assumed to be distributed as beta, then the beta-binomial model results:

$Y_{i.}|p_i \sim binomial(n_i, p_i)$ and $p_i \sim beta(a, b)$ where $Y_{i.} = \sum_j Y_{ij}$

$$Pr(Y_{i.} = y) = \binom{n_i}{y} \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) \left(\frac{\Gamma(y+a)\Gamma(n_i-y+b)}{\Gamma(a+b+n_i)}\right). \quad (2.5)$$

The marginal mean from the beta-binomial model is $E(Y_{ij}) = E(E(Y_{ij}|p_i)) = \frac{a}{a+b} = \mu_i$ and $corr(Y_{ij}, Y_{ik}) = \frac{1}{a+b+1} = \delta_i$ is the within-cluster correlation. The binomial model is a special case of the beta-binomial model when $\delta_i = 0$ (Paul, 1982).

Standard likelihood techniques can be used when the marginal mean and within-cluster correlation are of interest. The beta-binomial model assumes that the ordering of the index j is random, i.e. that the within-cluster correlation is the same for all pairs of observations within a cluster.

The relationship between covariates and the marginal mean risk and/or the within-cluster correlation may be of interest, where (as in WCR) the mean risk parameter corresponds to the risk for a random pup drawn from a random litter. Originally, models for the marginal mean and the within-cluster correlation parameters could only involve cluster-specific covariates, where each parameter is a function of the beta parameters, a and b (Prentice, 1986). (Another maximum likelihood (ML) method that only accommodates cluster-specific covariates is the Zeger, Liang, and Self (1985) method for longitudinal data, discussed later in this section.) Rosner (1984), Qu, Williams, Beck, and Goormastic (1987), and Connolly and Liang (1990) have extended the beta-binomial model to include observation-specific, as well as cluster-specific covariates. Similarly, Zeger and Liang (1986) extended their earlier model (Zeger, Liang, and Self, 1985) to include observation-specific covariates as well. These extended models are included in the discussion of response-conditional models (Section 2.5) and quasi-likelihood methods (Section 2.3.3), respectively.

Other litter-effects problem approaches include a "heterogeneity adjustment of variance" (Finney, 1973; Segreti and Munson, 1981), the correlated-binomial model (Kupper and Haseman, 1978; Altham, 1978), the multiplicative binomial model (Altham, 1978), the jackknife method (Gladen, 1979), and a quasi-likelihood method (Williams, 1982). Finney (1973) proposed a simple approach that uses a logistic model with a "heterogeneity adjustment of variance". A significant test of heterogeneity, χ^2, with degrees of freedom, df, indicates extra-binomial variation. Therefore, to account for this variation without a complicated model, Finney (1973) suggested multiplying the variances by the heterogeneity factor, $\frac{\chi^2}{df}$.

Two competitors of the beta-binomial model are the correlated-binomial (Kupper and Haseman, 1978; Altham, 1978) and the multiplicative binomial models (Altham, 1978). Instead of modeling the heterogeneity among the litters like the beta-binomial, the correlated-binomial explicitly models the correlation among pups within a litter (Kupper and Haseman, 1978; Altham, 1978; Paul, 1982). The multiplicative binomial model is a two parameter generalization of the binomial model (Altham, 1978; Paul, 1982). Paul (1982) showed by simulation that both the correlated-binomial and the multiplicative binomial models are less sensitive to departures from the binomial than is the beta-binomial model. A cited advantage of the correlated-binomial and the multiplicative binomial models over the beta-binomial model is their ability to accommodate both positive and negative correlation (Kupper and Haseman, 1978; Altham, 1978; Haseman and Kupper, 1979; Paul, 1982). However, by re-parameterizing the beta-binomial model, Prentice (1986) showed that negative correlations in a neighborhood around zero could mathematically be accommodated in an extended beta-binomial model.

Gladen (1979) originally developed the jackknife method to reduce the bias of estimators of proportions from litter-effects data and to produce estimates of the standard error for the estimators. A pseudo t-test can be obtained by the jackknife method, which

is easy to compute and has been shown by simulation to perform as well as a beta-binomial likelihood ratio test (Gladen, 1979; Paul, 1982).

Williams (1982) proposed a quasi-likelihood approach for analyzing litter-effects data. An advantage of the quasi-likelihood approach is that a specific form of the distribution of the outcomes is not required, but only the relationship between the first two moments is assumed known. Quasi-likelihoods will be discussed in more detail in Section 2.3.3.

The Lipsitz, Laird and Harrington (1990) ML method is for paired, structured binary data. It allows for both cluster-specific and observation-specific covariates but only for clusters of size two, $n_i = 2$. This is a severe limitation since litter-effects type data or longitudinal data usually have cluster sizes that are greater than two.

Maximum likelihood estimation (MLE) is used for the paired data method since the likelihood is explicitly specified by the multinomial distribution corresponding to the four possible paired outcomes. The marginal probabilities are then modeled logistically as a function of the covariates. The odds ratio, relative risk, or the correlation coefficient can represent the dependency between any two observations. The choice of dependency parameter depends on which one is easiest to interpret in a particular analysis. The dependency parameter can be modeled as a function of the covariates with additional parameters, to increase the flexibility of the model. This is true of the within-cluster correlation parameter in the beta-binomial model as well (Prentice, 1986). A nice feature of the paired data model, which is not shared by the beta-binomial or the Zeger, et al. (1985) models, is the ability to choose the type of dependency parameter for an analysis.

Zeger, Liang, and Self (1985) extended logistic regression to the case of repeated binary outcomes, such as in a longitudinal setting. However, their method can only incorporate cluster-specific covariates and two specific structures for the within-cluster correlation. As noted previously, the beta-binomial model also has the cluster-specific covariate constraint.

In the Zeger, et al. (1985) model, the working assumptions are that the observations within a cluster are either independent or from a first-order Markov chain, $\rho = corr(Y_{ij}, Y_{i,j-1})$. Since the exact joint distributions for both of the dependence assumptions can be explicitly written, MLE can be used. The marginal expectations are modeled as functions of the covariates. The parameter estimates, $\underset{\sim}{\widehat{\beta}}$, are obtained by MLE. Zeger, et al. (1985) showed that as $I \rightarrow \infty$, $I^{1/2}(\underset{\sim}{\widehat{\beta}} - \underset{\sim}{\beta})$ is distributed asymptotically multivariate normal (MVN). This approach has limited use since the exact likelihood must be specified and only cluster-specific covariates can be used.

The parameters from each of these ML methods are interpreted as population-averaged effects. Also, each was developed for a specific type of data. Typically, the beta-binomial model is used with litter-effects type data; the Lipsitz, et al. (1990) model is used for paired data, such as eyes or ears; and the Zeger, et al. (1985) models are used in conjunction with data observed over time. These distinctions, along with others noted, should provide some direction in choosing an appropriate ML analysis method for a particular situation.

2.3.3 Quasi-likelihood Methods

The generalized estimating equation (GEE) approach (Liang and Zeger, 1986; and Zeger and Liang, 1986) is a widely applicable method for analyzing clustered binary data by a marginal model. It allows observation-specific covariates, as well as cluster-specific covariates, which is an improvement over the beta-binomial and the Zeger, et al. (1985) models. In addition, the GEE approach does not require the exact specification of the likelihood, which is required for ML methods. The GEE approach is highly dependent upon generalized linear model (GLM) (McCullagh and Nelder, 1983) and quasi-likelihood theory (Wedderburn, 1974).

Quasi-likelihood theory states that only the relationship between the mean and the variance needs to be specified rather than the actual distribution. An estimating equation is the quasi-likelihood generalization of the score equation formed by taking the derivative with respect to the parameter of interest of the quasi-likelihood function. One solves the estimating equation to obtain the maximum quasi-likelihood estimates (MQLE). Wedderburn (1974) developed quasi-likelihood theory for independent data, and Liang and Zeger (1986) extended it for use with dependent data. The GEE approach provides a way to analyze clustered binary data in situations where the likelihood cannot be specified easily. The GEE method extends the Zeger, et al. (1985) approach to allow observation-specific, as well as cluster-specific covariates and more complicated dependency structures.

The GEE approach is based on a marginal model; the parameters from this model are interpretable as population-averaged effects. GEE models account for the within-cluster correlation by assuming a user-specified working correlation matrix. However, the regression estimates and their estimated standard errors are consistent even if the dependency is mis-specified, as long as the model for the marginal mean is specified correctly. The efficiency of the parameter estimates increases if the chosen working correlation structure is the true correlation structure.

Assume that $\underset{\sim}{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$ is a (n_ix 1) vector of outcomes and $\underset{\sim}{X}_i = (\underset{\sim}{x}_{i1}, \ldots, \underset{\sim}{x}_{in_i})^T$ is a (n_ix p) matrix of covariates for cluster i, $i = 1, \ldots, I$. Assume the link and variance functions are written as in (2.1). The mean of an observation is $E(Y_{ij}) = \mu_{ij} = h^{-1}(\eta_{ij})$, where $\eta_{ij} = \underset{\sim}{x}_{ij}\underset{\sim}{\beta}$ represents the linear predictor, and the variance of an observation is $var(Y_{ij}) = g(\mu_{ij}) \cdot \phi$.

The GEE approach accounts for the within-cluster correlation by inserting a working correlation matrix, $R_i(\underset{\sim}{\alpha})$, into the covariance formula

$$\underset{\sim}{V}_i = \underset{\sim}{A}_i^{1/2} R_i(\underset{\sim}{\alpha}) \underset{\sim}{A}_i^{1/2} / \phi \tag{2.6}$$

where $\underset{\sim}{A}_i = \text{diag}(g(\mu_{ij}))$, $R_i(\underset{\sim}{\alpha})$ is $(n_i \text{x } n_i)$, and $\underset{\sim}{\alpha}$ $(s\text{x}1)$ is a vector of parameters that completely specifies $R_i(\underset{\sim}{\alpha})$ (Liang and Zeger, 1986). The generalized estimating equation replaces the score equation and is written as

$$\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{S}_i = \underset{\sim}{0} \tag{2.7}$$

where $\underset{\sim}{D}_i = \frac{\partial \mu_i}{\partial \underset{\sim}{\beta}}$ and $\underset{\sim}{S}_i = \underset{\sim}{Y}_i - \underset{\sim}{\mu}_i$. Note that the $\underset{\sim}{V}_i$ may be of varying dimension. For normally distributed data, (2.7) reduces to the least squares normal equations (Zeger, 1988). The parameter estimates are obtained by alternating between iteratively reweighted least squares (IRLS) solution of (2.7) and moment estimation of the correlation parameters, $\underset{\sim}{\alpha}$, and the scale parameter, ϕ. A common approach for estimating $\underset{\sim}{\alpha}$ and ϕ at each iteration uses functions of the Pearson residuals (Liang and Zeger, 1986). The assumed dependency structure in the model determines the exact form of the residual formulas.

Another approach for estimating the correlation parameters, $\underset{\sim}{\alpha}$, is the extended GEE approach (Prentice, 1988; Zhao and Prentice, 1990; Prentice and Zhao, 1991). Although the original GEE approach (Liang and Zeger, 1986; Zeger and Liang, 1986) provides consistent estimates when the dependency structure is not correctly specified, the efficiency of the GEE estimator may be low. To increase the efficiency of the GEE regression estimators, initially Prentice (1988) and later Zhao and Prentice (1990) proposed an extension of the original GEE method to jointly estimate the regression and correlation parameters. In addition to the original generalized estimating equations (2.7), a second set of estimating equations for the parameters in the user-specified working correlation matrix are specified. The regression and correlation parameters estimated by

the extended GEE approach are the same parameters in the ordinary GEE approach, namely $\underset{\sim}{\beta}$ and $\underset{\sim}{\alpha}$. The two sets of estimating equations are parameterized in terms of the marginal mean and the pairwise covariances for the estimation of the regression and correlation parameters.

Lipsitz, Laird, and Harrington (1991) advocate modifying the extended GEE method by parameterizing the covariance matrix in terms of pairwise odds-ratios instead of pairwise correlations. They found the odds-ratio parameterization to be slightly more efficient than the correlation parameterization, although the interpretation of the regression parameters and asymptotic properties are the same as the extended GEE method. Cessie and van Houwelingen (1994) also proposed a pairwise odds-ratio parameterization. The extended GEE approach allows simultaneous estimation of both the regression and correlation parameters. The joint score estimating equation from the extended GEE method is very similar to the original GEE (see Zhao and Prentice, 1990 or Prentice and Zhao, 1991 for more details).

The estimated regression parameters from original GEE are asymptotically MVN with mean of zero and covariance matrix

$$\underset{\sim}{V}_{GEE} = \lim_{I\to\infty} I(\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1} \{\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} cov(\underset{\sim}{Y}_i) \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i\} (\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1}. \qquad (2.8)$$

The asymptotic covariance matrix, (2.8), requires consistency of the estimated correlation parameters, $\hat{\underset{\sim}{\alpha}}$, and the estimated scale parameter, $\hat{\phi}$. However, in practice, these necessary conditions are rarely verified and routinely assumed to hold. This ignorance may lead to situations where (2.8) is not the asymptotic covariance matrix.

The covariance is robustly estimated by substituting $\underset{\sim}{S}_i \underset{\sim}{S}_i^T$ for $cov(\underset{\sim}{Y}_i)$ and the estimates of β, α, and ϕ into (2.8). When $R(\alpha)$ is the "true" correlation matrix, then the asymptotic covariance matrix in (2.8) simplifies to $\lim_{I\to\infty} I(\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1}$. The inner

term of (2.8) provides the robustness to mis-specification of the dependency by using the covariance matrix from the data to estimate the within-cluster dependency, whereas the outer terms rely on the working correlation matrix.

Crowder (1995) found that when the user-specified working correlation matrix is inappropriate, there may be no solution for the correlation parameters (Crowder, 1995). When this happens, the asymptotic properties of the regression and correlation parameters break down and consistency cannot be claimed. To avoid this potential problem, Crowder (1995) suggests only using estimating equations that have a guaranteed solution.

The estimated regression and correlation parameters from the extended GEE are also asymptotically MVN with zero mean. The robust covariance matrix for the extended GEE is virtually identical to (2.8), except the covariance matrix includes additional variance components from the second estimating equation. If the regression and correlation estimating equations from the extended GEE are "independent" in the sense that there are two distinct sets of equations, one involving only the regression parameters, $\underset{\sim}{\beta}$, and the other involving only the correlation parameters, $\underset{\sim}{\alpha}$, then the regression estimates and corresponding standard errors will be consistent, even if the dependency is not correctly specified (Zhao and Prentice, 1990). However, if they are not "independent" of each other, then consistency occurs only when the mean and the covariances are modeled correctly. Since the correct form of the dependency is often not known, there is an increased risk with the extended GEE, compared with ordinary GEE, of biased estimates when the regression and correlation parameters are estimated simultaneously. Whether the risk of biased regression parameters outweighs the quest for more efficient estimates of the regression parameters depends upon the prior knowledge, the judgment of the analyst, and the context. Liang, Zeger, and Qaqish (1992) suggest using ordinary GEE when the number of clusters is large and the correlation parameters

are considered nuisances. If there are only a few clusters and/or the correlation parameters are of primary interest, then they recommend the extended GEE method.

Use of either GEE method for large cluster sizes with relatively few clusters ($n_i > I$)is not supported by the asymptotic theory, which depends on the number of clusters being large. Therefore, the robust variance (2.8) can perform poorly in an analysis with a few clusters of large size.

Since data sets are routinely not complete, it is important to understand how missing data patterns affect the GEE approaches. The semi-parametric approach of GEE requires any missing data to be missing completely at random (Rubin, 1976) to retain its robustness properties (Liang and Zeger, 1986). Missing completely at random (MCAR) means that the missingness is independent of the outcome and covariate vectors (Little and Rubin, 1987). The GEE method will not provide valid inference if the missing data are merely missing at random (Liang and Zeger, 1986; Fitzmaurice, Laird, and Lipsitz 1994). Missing at random (MAR) means that conditional on the covariate vector, the probability that the outcome is missing is independent of the outcome (Little and Rubin, 1987).

Other variations on the GEE and extended GEE approaches have been proposed. Bieler and Williams (1995) implemented an existing approach that has been used historically for complex surveys. Since complex surveys deal with correlated outcomes, the application of the survey procedure may be appropriate for other types of correlated data. The close connection between the survey and the GEE approaches has been demonstrated by Bieler and Williams (1995), LaVange, et al. (1994), and Rotnitzky and Jewell (1990). They show that the main difference is in the divisor of the variance (survey approach: $I - 1$ vs. GEE approach: I). The survey approach can be implemented by the software package SUDAAN (Research Triangle Institute, 1995).

Non-convergence can be a problem for both of the GEE methods. Lipsitz, Fitzmaurice, Orav, and Laird (1994) propose a one-step estimator to overcome

convergence problems with ordinary or extended GEE methods, such as with small sample sizes (see Lipsitz, Fitzmaurice, Orav, and Laird (1994) for more details). "Alternating logistic regression" was suggested by Carey, Zeger, and Diggle (1993) to obtain more efficient parameter estimates when the computations for the extended GEE (Zhao and Prentice, 1990) become unfeasible, such as with large cluster sizes (see Carey, Zeger, and Diggle (1993) for more details).

Another method, similar to GEE, is that of Wei and Stram (1988). They propose analyzing clustered binary data that arise as longitudinal observations made at a specified set of time points. Estimates of the regression parameters are obtained separately for each time point. Asymptotic joint normality of the parameter estimates, $(\widehat{\beta}_1, \ldots, \widehat{\beta}_T)$ is claimed by the Multivariate Central Limit Theorem. This approach is limited because longitudinal data sets rarely have the same set of time points for each cluster. The issues relating to the original GEE method also are relevant for this method. Stram, Wei and Ware (1988) extended this approach to ordinal data.

2.3.4 Bootstrap Method

An interesting marginal approach developed by Moulton and Zeger (1989) incorporates the bootstrap method into the analysis of longitudinal binary data with common time points. Similar to the Wei and Stram (1988) method, the bootstrap method analyzes the data separately at each time point to obtain an estimate of the parameters of interest at time t, say $\widehat{\underset{\sim}{\beta}}_t$, $t = 1, \ldots, T$ via iteratively weighted least squares (IWLS). In order to account for the within-cluster correlation, the data vectors $(y_{i1}, \ldots, y_{iT}; \underset{\sim}{x}_{i1}, \ldots, \underset{\sim}{x}_{iT})$ are sampled with replacement to obtain the b^{th} bootstrap replicated data set. For each of the $b = 1, \ldots, B$ bootstrap replications, the original value of $\widehat{\underset{\sim}{\beta}}$ is used as the starting value for the bootstrap data which is analyzed separately at each time, t. One iteration of a IWLS calculation is performed to avoid problems with

boundary values and to get a new value of the vector of parameters of interest, $\underset{\sim}{\hat{\beta}}^{*b}$. This process is repeated. This bootstrap approach provides an approximation of the distribution of $\underset{\sim}{\hat{\beta}}$. The parameter estimates may be combined across time. The variance for the one-step bootstrap coefficient is identical to the robust variance estimator discussed by Royall (1986). When there are the same number of observations at each time point, this method is asymptotically equivalent to the independence estimator of Zeger, et al. (1985) (Moulton and Zeger, 1989)

This approach is limited, like the Wei and Stram (1988) method, because of the type of data required. Also, it is not commonly used, unless the parameter at each time point is of interest. In addition, it is very computationally intensive.

2.4 Random-effects Models

2.4.1 Introduction

Random-effects or subject-specific models provide an alternative method for analyzing clustered binary data. Random-effects models assume there is heterogeneity across clusters in the population that is responsible for the within-cluster correlation. The within-cluster dependency is often modeled implicitly by assuming a parametric model for the among-cluster heterogeneity. However, for a multiplicative model there are non-parametric random-effects models, such as conditional logistic regression. Random-effects models for clustered binary data are not as widely used as marginal models because they often require intense computations, an expectation and maximization (EM) algorithm, and iterative methods.

2.4.2 Conditional Logistic Regression

Conditional logistic regression is a way to avoid bias when the number of parameters increases with the sample size, as occurs with many small clusters when one

attempts to estimate cluster-specific parameters (Breslow and Day, 1980; Hosmer and Lemeshow, 1989; Stokes, Davis, and Koch, 1995). It is a non-parametric approach because the cluster-specific parameter, which is viewed as a nuisance parameter, is conditioned out. Therefore, it is not necessary to assume a distribution for the random effect. Let the model be as (2.2) with a logit link,

$$Pr(y_{ij}) = \frac{exp(\alpha_i+\beta x_{ij}+\underset{\sim}{\gamma}^T \underset{\sim}{z}_{ij})}{1+exp(\alpha_i+\beta x_{ij}+\underset{\sim}{\gamma}^T \underset{\sim}{z}_{ij})} \tag{2.9}$$

where α_i is the effect of the cluster i, β is the parameter for x_{ij} (the effect of interest), and $\underset{\sim}{\gamma}$ is the parameter vector for other covariates, $\underset{\sim}{z}_{ij}$.

Consider the special case where the clusters are all of size two. Using conditional probabilities based on the pairs that provide information, the discordant pairs, for the case where $n_i = 2$,

$$Pr(y_{i1} = 1, y_{i2} = 0 |\ y_{i1} = 1, y_{i2} = 0 \text{ or } y_{i1} = 0, y_{i2} = 1, \underset{\sim}{x}, \underset{\sim}{z}) =$$

$$\frac{Pr(y_{i1}=1|\underset{\sim}{x},\underset{\sim}{z})Pr(y_{i2}=0|\underset{\sim}{x},\underset{\sim}{z})}{Pr(y_{i1}=1|\underset{\sim}{x},\underset{\sim}{z})Pr(y_{i2}=0|\underset{\sim}{x},\underset{\sim}{z})+Pr(y_{i1}=0|\underset{\sim}{x},\underset{\sim}{z})Pr(y_{i2}=1|\underset{\sim}{x},\underset{\sim}{z})}$$

the conditional logistic regression model reduces to the familiar logistic regression model

$$\frac{exp(\beta+\underset{\sim}{\gamma}^T(\underset{\sim}{z}_{i1}-\underset{\sim}{z}_{i2}))}{1+exp(\beta+\underset{\sim}{\gamma}^T(\underset{\sim}{z}_{i1}-\underset{\sim}{z}_{i2}))}. \tag{2.10}$$

(Stokes, Davis, and Koch, 1995). Equation (2.10) looks like (2.9) except that β, the effect of interest, is now the intercept, and the other covariates, $\underset{\sim}{z}_{ij}$, are the differences between the covariates within a pair, $(\underset{\sim}{z}_{i1} - \underset{\sim}{z}_{i2})$. Conditional logistic regression on pairs only requires minor modifications of a data set and can be fit by any logistic

regression software. With cluster sizes greater than two, specialized software is required and this approach becomes computationally intensive when the cluster size is moderate. The conditional likelihood approach is a way to estimate the parameters of interest without bias, while avoiding the need to estimate the random effects.

2.4.3 Methods with Gaussian Random Effects

A majority of the random-effects models in the literature assume that the random effects are normally distributed. This statistically convenient assumption is typically hard to verify.

Random-effects models are sometimes referred to as two-stage or hierarchical models. Stiratelli, Laird, and Ware (1984), Zeger and Karim (1991), Breslow and Clayton (1993), and Wolfinger and O'Connell (1993) adapt the general random-effects method of Laird and Ware (1982) to analyze clustered binary data. Here a general logistic-linear mixed model is assumed. The probability of a positive response for cluster i at time j is represented by $p_{ij} = \Pr(y_{ij} = 1)$, $\lambda_{ij} = \text{logit}(p_{ij})$, $\underset{\sim}{X}_i$ is a matrix of covariates, which may consist of observation-specific and cluster-specific covariates, and $\underset{\sim}{Z}_i$ is often a subset of $\underset{\sim}{X}_i$. The two stages of the model for clustered binary data are

$$\text{Stage 1:} \qquad \underset{\sim}{\lambda}_i = \underset{\sim}{X}_i \underset{\sim}{\alpha} + \underset{\sim}{Z}_i \underset{\sim}{b}_i \tag{2.11}$$

$$\text{Stage 2:} \qquad \underset{\sim}{b}_i \sim MVN(\underset{\sim}{0}, \underset{\sim}{B}) \tag{2.12}$$

as described by Laird and Ware (1982) except that in this case, $\underset{\sim}{\lambda}_i$ represents the vector of logits for cluster i. Sometimes a flat or non-informative prior is assumed for $\underset{\sim}{\alpha}$, such as $\underset{\sim}{\alpha} \sim MVN(\underset{\sim}{0}, \underset{\sim}{\Gamma})$ and $\underset{\sim}{\Gamma}^{-1} \to 0$. The empirical Bayes framework is applied when the fixed effects have prior distributions. This is the approach advocated by Stiratelli, et al. (1984) and Zeger and Karim (1993).

Similarly, Breslow and Clayton (1993) and Wolfinger and O'Connell (1993), outline two-stage models. However, instead of using the empirical Bayes with an expectation/maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977), Breslow and Clayton propose a penalized quasi-likelihood (PQL) method that uses Laplace's integral approximation (Barndorff-Nielsen and Cox 1989; Tierney and Kadane, 1986). The approach of Wolfinger and O'Connell is identical to the PQL approach except pseudo-likelihood (PL) is used. Wolfinger and O'Connell show that PQL is a special case of PL when the scale parameter, ϕ, equals one. The PQL, PL, and empirical Bayes approaches result in the same set of equations for the fixed-effect and random-effect parameters. However, they differ in how the marginal (quasi- or pseudo-) likelihood $L(\underset{\sim}{\alpha}, \underset{\sim}{B})$ (the likelihood after the random effect, $\underset{\sim}{b}_i$, has been integrated out) is approximated. Since the PL and PQL approaches are virtually identical, I will only describe the PQL approach.

Stiratelli, et al. (1984) use empirical Bayes methods to estimate the random-effect parameters, $\underset{\sim}{\alpha}$ and $\underset{\sim}{b}_i$, and their covariances, $\underset{\sim}{\Sigma}_{\alpha}$ and $\underset{\sim}{\Sigma}_{b_i}$. The modes of the posterior distributions are used to approximate the posterior means in estimating $\underset{\sim}{\alpha}$ and $\underset{\sim}{b}_i$.

The Stiratelli, et al. approach obtains all of the estimates and predicted values from the EM algorithm. The derivatives of the marginal likelihood, $L(\underset{\sim}{B}, \underset{\sim}{\Gamma}^{-1} = 0)$, (the marginal likelihood after both $\underset{\sim}{\alpha}$ and $\underset{\sim}{b}_i$ have been integrated out) are used to estimate the covariance parameters in $\underset{\sim}{B}$. The EM algorithm is an iterative method that alternates between two steps. The two steps are the "estimation" step called the E-step, and the "maximization" step called the M-step (see Dempster, Laird, and Rubin, 1977 and Laird and Ware, 1982 for more details on the EM algorithm). The original development of the EM algorithm was for incomplete data settings. However, Laird and Ware (1982) extended the EM algorithm to random-effects estimation, where the random effects are assumed to be missing data. For clustered binary data, Newton-Raphson iteration is

required at each of the E-steps, making this method quite computationally intense (Stiratelli, et al., 1984). Also, the EM algorithm often converges slowly and can tend to a local instead of a global maximum (Laird and Ware, 1982).

The PQL approach results in the same set of equations as the empirical Bayes method, but it uses the integral approximation of Laplace. The estimates of $\underset{\sim}{\alpha}$, $\underset{\sim}{b}_i$, $\underset{\sim}{\Sigma}_\alpha$, and $\underset{\sim}{\Sigma}_{b_i}$ are obtained by Fisher scoring (Green, 1987). The PQL gets the estimates of the parameters of the random effects distribution, $\underset{\sim}{\hat{B}}$, by maximizing the modified profile likelihood, which results from imputing the estimates of the fixed effects and the predicted values of the random effects into the penalized quasi-likelihood equation. Both approaches provide estimates of $\underset{\sim}{\alpha}$, $\underset{\sim}{\Sigma}_\alpha$, $\underset{\sim}{\Sigma}_{b_i}$, $\underset{\sim}{B}$, and predicted values of $\underset{\sim}{b}_i$.

Zeger and Karim (1991) use the GLM models of (McCullagh and Nelder, 1989; Nelder and Wedderburn, 1972) and the Bayesian framework of Stiratelli, et al. (1984). However, they adopt Gibbs sampling, a Monte Carlo method, (Gelfand and Smith, 1990; Geman and Geman, 1984) to estimate the posterior distributions. This approach is advantageous when there is more than just a random intercept in the model and for specifying a number of distributions other than the normal for the random effects.

Gilmour, Anderson, and Rae (1985) propose an "expectation" method for the Gaussian probit model. Their method maximizes the likelihood with respect to the fixed effects, while taking the expected value over the random-effects (Gilmour, Anderson, and Rae, 1985). The expectation method is contrasted with the "joint-maximization" method of Gianola and Foulley (1982, 1983) and Harville and Mee (1984), who propose maximizing the likelihood for both the fixed and random effects in a Bayesian framework.

Anderson and Aitkin (1985) use a variance component model for interviewer variability. This method assumes that the binary outcomes are realizations of underlying normal variables. A multivariate normally distributed random effect for each cluster is specified in a logistic model. Their method is an iterative maximum likelihood estimation

method for finding the fixed effects and variance parameters from the random effects distribution. The integral over the random-effects distribution is approximated by a summation over a finite number of points. The ML estimators are obtained by iteratively weighted least squares. This method allows observation-specific covariates to be included in the model (Anderson and Aitkin, 1985). However, unlike the empirical Bayes, PQL, PL, and Gibbs sampling approaches, this method does not provide predicted values of the random effects.

2.4.4 Method With Non-Gaussian Random Effects

Even though the Zeger and Karim (1991) approach can have non-Gaussian random effects, Conaway (1990) proposed a subject-specific model that explicitly assumes a log-gamma distribution for the random effects with the log-log link. This approach does not require numerical integration to obtain the marginal likelihood, $L(\underset{\sim}{\alpha}, \underset{\sim}{\beta})$. The log-gamma distribution allows many different shapes for the random-effects distribution, some of which are very similar to the normal distribution. However, the parameters from this approach are difficult to interpret in general and are not log odds-ratios.

Conaway (1990) also suggests an interesting mix of models. He proposes a conditional random-effects model (see Section 2.5 for conditional models), where the heterogeneity in the population is modeled by the log-gamma distribution and the previous outcomes are used as covariates in the model. This is similar to the regressive logistic models of Bonney (1987), but with the addition of random effects in the model.

2.5 Conditional Models

2.5.1 Introduction

Conditional models account for the within-cluster correlation by conditioning on the past outcomes. A Markov assumption is commonly made in conditional models,

where the current response only depends upon the past through a limited number of prior responses. Conditional models can be fit by standard software with the past outcomes as covariates

$$h(\mu_{ij}^c) = x_{ij}^T \beta^c + \sum_k \alpha_k^c f_k(Y_{i1}, \ldots, Y_{ij-1}), \tag{2.13}$$

where f_k, $k = 1, \ldots, \nu$, are known functions of the past outcomes (Bonney, 1987; Zeger and Liang, 1992).

2.5.2 Regressive Logistic Model

The conditional model is a simple way to specify the joint probability, which itself may be too complicated to work with. This is similar to the structure of a log-linear model, which is discussed by Bonney (1987) and Liang, Zeger, and Qaqish (1992). The conditional approach for clustered binary data is exemplified well by Bonney (1987). He expresses the joint probability as the product of independent conditional probabilities

$$\begin{aligned} Pr(\underset{\sim}{Y}|\underset{\sim}{X}) &= Pr(Y_1, Y_2, \ldots, Y_n|\underset{\sim}{X}) \\ &= Pr(Y_1|\underset{\sim}{X})Pr(Y_2|Y_1, \underset{\sim}{X}) \cdots Pr(Y_n|Y_1, Y_2, \ldots, Y_{n-1}, \underset{\sim}{X}). \end{aligned} \tag{2.14}$$

Each of the conditional probabilities is then modeled logistically as a function of covariates and the previous outcomes

$$Pr(Y_i|Y_1, Y_2, \ldots, Y_{i-1}, \underset{\sim}{X}) = \frac{exp(\omega_i Y_i)}{1+exp(\omega_i)}, \tag{2.15}$$

where ω_i is a linear function of the covariates and the previous outcomes. Many different dependency structures can be accommodated by this method.

This conditional method, as previously mentioned, is only useful for data with a natural ordering, such as longitudinal data. The model is highly dependent upon the specific ordering of the data because different orderings will normally produce different estimates (Bonney, 1987).

2.5.3 Response-conditional Model

The response-conditional model is a slight variation of the conditional model. The response-conditional model depends on *all* of the other outcomes instead of just the *previous* outcomes. A class of conditional logistic models, termed response-conditional models, has been studied by many statisticians, including Rosner (1984), Prentice (1988), Qu, et al (1987), and Connolly and Liang (1990).

Rosner (1984) proposed analyzing ophthalmologic data by assuming that the cluster-specific baseline risk of disease in an eye is distributed beta(a, b) across clusters. The two eyes have independent Bernoulli outcomes, conditional on the cluster risk. Using the ophthalmologic data, there are four possible disease states for the outcomes of the eyes: $(++, +-, -+, --)$. Rosner shows that the probabilities associated with the disease states are functions of the parameters of the beta distribution, a and b. The logit of the conditional probabilities $p_R(p_L)$, probability of the right (left) eye being diseased given the outcome of the left (right) eye, is assumed to be linear in the covariates and the outcome of the other eye

$$ln\left(\frac{p_R}{1-p_R}\right) = \alpha_1 + (\alpha_2 - 2\alpha_1)z_L + \underset{\sim}{\beta}\,\underset{\sim}{x}^{(0)} + \underset{\sim}{\gamma}\,\underset{\sim}{x}^{(1)} \tag{2.16}$$

$$ln\left(\frac{p_L}{1-p_L}\right) = \alpha_1 + (\alpha_2 - 2\alpha_1)z_R + \underset{\sim}{\beta}\,\underset{\sim}{x}^{(0)} + \underset{\sim}{\gamma}\,\underset{\sim}{x}^{(2)} \tag{2.17}$$

where $\alpha_1 = ln\left(\frac{a}{b+1}\right), \alpha_2 = ln\left(\frac{(a+1)a}{(b+1)b}\right)$, z_R (z_L) is an indicator that the right (left) eye is diseased, $\underset{\sim}{x}^{(0)}$ is the vector of cluster-specific covariates, and $\underset{\sim}{x}^{(1)}$ $(\underset{\sim}{x}^{(2)})$ is the vector of the right (left) observation-specific covariates (Rosner, 1984). This approach extends the beta-binomial model to included observation-specific covariates. A polytomous logistic regression model follows directly from (2.16) and (2.17)

$$Pr(j|\underset{\sim}{x}_i) = \frac{exp(\underset{\sim}{\theta}_j \underset{\sim}{x}_i^*)}{\sum_{k=1}^{4} exp(\underset{\sim}{\theta}_k \underset{\sim}{x}_i^*)} \tag{2.18}$$

where $\underset{\sim}{x}_i^* = \{1, (\underset{\sim}{x}_i^{(0)})^T, (\underset{\sim}{x}_i^{(1)})^T, (\underset{\sim}{x}_i^{(2)})^T\}$, $j = 1, 2, 3,$ or 4 correspond to the four disease states, and θ is the matrix of parameters

$$\underset{\sim}{\theta} = \begin{bmatrix} \underset{\sim}{\theta}_1 \\ \underset{\sim}{\theta}_2 \\ \underset{\sim}{\theta}_3 \\ \underset{\sim}{\theta}_4 \end{bmatrix} \equiv \begin{bmatrix} \alpha_2 & 2\underset{\sim}{\beta} & \underset{\sim}{\gamma} & \underset{\sim}{\gamma} \\ \alpha_2 & \underset{\sim}{\beta} & \underset{\sim}{\gamma} & \underset{\sim}{0} \\ \alpha_1 & \underset{\sim}{\beta} & \underset{\sim}{0} & \underset{\sim}{\gamma} \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix} \tag{2.19}$$

(Rosner, 1984). The equations of (2.18) can be solved by maximum likelihood with the Newton-Raphson method. The interpretation of the parameters is conditional on the outcome of the other eye. Rosner extends the polytomous logistic regression model to unequal cluster sizes and clusters of size greater than two by conditioning on the sum of the other outcomes and the number of outcomes (Rosner, 1984).

Connolly and Liang (1988) also proposed a general class of conditional logistic models that uses an estimating equation approach for analyzing clustered binary data, which includes the Rosner model as a special case. Connolly and Liang state that equations (2.16) and (2.17) are identical to those developed by Qu, et al. (1987). Like Connolly and Liang (1988), Qu, et al. (1987) generalized the Rosner (1984) approach. Qu, et al. show the derivations of the response-conditional model in greater detail than

Rosner (1984). Also, Qu, et al. allow for negative correlation by using the Polya-Eggenberger distribution instead of the beta-binomial distribution to model the correlation structure (Eggenberger and Polya, 1923; Johnson and Kotz, 1969; Qu, et al., 1987).

A disconcerting feature of the Rosner model (the Rosner, Connolly and Liang, and Qu, et al. models) is that it reduces to the beta-binomial distribution only when all of the covariates are zero. Therefore, special consideration is given to the zero level of the covariates. Prentice (1988) notes that a different parameterization would allow the model with the zero level and the non-zero level of the covariates to play symmetric roles. However, all of the problems associated with conditional models still exist with this re-parameterization.

The Rosner model and the Prentice model (1986) both use the beta-binomial distribution. The main difference between the two approaches is that the Rosner model can use observation-specific covariates whereas the Prentice model can only use cluster-specific covariates. Prentice models the joint distribution of the observations of a cluster by the beta-binomial model, so one conditions on the sum of the observations. The beta-binomial parameters, the mean and the correlation, are then modeled as a function of the covariates. However, this only allows cluster-specific covariates since the joint distribution of the observations (sum) is used. Rosner (1984), Qu, et al. (1987) and Connolly and Liang (1988) use the same model as Prentice without covariates, but incorporate the cluster-specific and observation-specific covariates into the joint distribution by modeling each conditional probability as a logistic function of the covariates. The parameters from all of these approaches (Rosner, Qu, et al., Prentice, and Connolly and Liang) derive interpretation from the conditional distribution.

2.6 Summary

There are numerous ways to analyze clustered binary data, most of which are laden with assumptions and intense computations. (For another review of on clustered

binary data methods, see Pendergast, et al., 1996.) The parameters from marginal models are interpreted as population-averaged effects. Although there is not an assumed distribution for the heterogeneity in the population, the variance formula accounts for the within-cluster correlation. Even though there are marginal models that explicitly model the association parameters, the marginal parameter is still the focus of population-averaged models. Marginal models are widely used because they are easy to implement and interpret.

If the change in a covariate within a cluster or subject-specific inference is of importance, then a random-effects model may be the more appropriate choice. Random-effects models look at how changes in the values of the covariates within a cluster affect the response within that cluster. Parametric random-effects models assume a specific distribution for the heterogeneity in the population, which is usually a normally-distributed random intercept term. The normal distribution is the traditional choice for the distribution of the random effect, but this choice is arbitrary. As shown by Conaway (1990), the log-gamma distribution can mimic the shape of the normal distribution and does not require numerical integration when used with the log-log link. However, the true distribution of the random effect is virtually impossible to determine. Random-effects models may not be used as often as marginal models because they require specialized software and are usually computationally intensive even with moderate cluster sizes (see Weinberg and Gladen, 1986 for a discussion of an exception).

Conditional models are another way to analyze clustered binary data when there is a natural ordering, such as longitudinal data. These models condition on the previous outcomes or all of the other outcomes in response-conditional models. Conditioning on other outcomes accounts for the within-cluster correlation, however, it may lessen the estimated effect of the covariates. The interpretation of the parameters from conditional models depends on the other outcomes and may vary with cluster size. Consequently, the parameter interpretation may not be invariant across studies with differing distributions of

cluster sizes. Table 2.1 displays a general comparison of marginal, random-effects, and conditional models.

Table 2.1. General Comparison of Characteristics of Models

TYPE OF MODEL	REQUIREMENTS	EASE OF FITTING	PARAMETER INTERPRETATION	SMALL SAMPLES
Marginal (GEE)	• User-specified working correlation matrix • Within-cluster correlation regarded as nuisance • Relationship between mean and variance can be specified	• Requires special software, available in SAS or S-plus • Problems with consistency of estimators can occur with user-specified working correlation matrix (Crowder, 1995)	• "Population-averaged"	• Approximations assume a large number of clusters • Convergence problems with small sample sizes • Estimation of precision and estimators may be biased
Conditional Logistic	• Dependency is secondary to heterogeneity • Distribution of heterogeneous risk is not of interest • Multiplicative model	• For paired data, can be fit by taking differences • For cluster size >2, requires special software, such as EGRET or Epicure. • Only uses clusters with variation in outcome and in covariates	• Subject-specific parameters conditioned out, so parameters interpreted as "within-cluster" effects	• Approximations assume a large number of clusters
Random Effects	• Dependency is secondary only to heterogeneity among clusters • Known distribution for random effect	• Requires special software, such as EGRET or macro in PROC MIXED in SAS • Computationally intense • Often, requires iterative methods that are slow to converge	• Parameters interpreted as average "within-cluster" effects • However, interpretation more like "population-averaged" effects when associated with a cluster-specific covariate	• Approximations assume a large number of clusters
Conditional	• Within-cluster correlation accounted for by conditioning on past outcomes • Parameter not dependent upon time	• Bonney (1987) model can be fit by logistic regression software with past outcomes as covariates • Rosner (1984) approach requires special software	• Interpretation dependent on the past outcomes • Only makes sense with longitudinal data	• Approximations assume a large number of clusters

WCR is an alternative method for analyzing clustered binary data. WCR is a cluster-based method that has a slightly different interpretation than the models previously discussed. WCR estimates the difference in response of a randomly selected observation with covariate x from a random cluster versus a randomly selected observation with covariate $x + 1$ from a random cluster. With non-informative cluster sizes, the WCR interpretation is identical to the interpretation from a population-averaged model. Unlike the GEE method, the WCR method does not require a user-specified working correlation matrix. (Although this choice does not matter asymptotically, for typical sample sizes it can have a big effect on the estimated variance.) When there are only cluster-specific covariates, the WCR method can be regarded as a non-parametric random-effects model. Since WCR is a marginal approach, there is not an assumed distribution for the heterogeneity among the clusters.

CHAPTER III

THEORY OF WITHIN CLUSTER RESAMPLING

3.1 Introduction

Marginal models are a common approach for analyzing clustered data when the dependency is treated as a nuisance. Traditionally, marginal models relate the outcome variable to the covariates without explicitly modeling the dependency structure. Within Cluster Resampling (WCR) is a new marginal analysis method for clustered data.

A cluster is defined as a group of dependent outcomes, where the cluster may be specific to an individual, an animal, etc. The dependent outcomes nested within a cluster are referred to as observations. Examples might include whether or not a particular pregnancy of a woman resulted in a live-birth or whether or not a pup from a litter has a birth-defect. The theory that follows is general so that it encompasses binary, categorical, and continuous data; however, the examples focus on clustered binary data.

WCR equally weights each cluster in an analysis, whereas other marginal methods, such as the generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986), the weight increases with the size of the cluster. Equally weighting each cluster in an analysis may be beneficial for analyzing litter-effects data where it may be inappropriate to give greater weight to dams who naturally produce larger litters or in a longitudinal study where patients with reason for concern may be seen more in the clinic.

In Section 3.2. the generalized linear model notation and the asymptotic normality of a standardized estimator is stated as the first proposition. The WCR sampling scheme is outlined and discussed in detail in Section 3.3. Two propositions regarding consistency and asymptotic normality are proven in Section 3.4. The fourth proposition concerning the proposed WCR variance estimator is established in Section 3.5. The results of analyses of two toxicological data sets using WCR, GEE, and ordinary logistic regression (OLR) are presented in Section 3.6. A summary is given in Section 3.7.

3.2 The Generalized Linear Model

First, the generalized linear model (GLM) notation will be introduced. Hence, the classical proof of the asymptotic normality of a standardized estimator will be given. The notation and proof are important background for the proof of the asymptotic normality of the standardized WCR estimator, which is found in Section 3.4.

The generalized linear model (GLM) (McCullagh and Nelder, 1983) was introduced as an extension of the classical linear model. A generalized linear marginal model can be written as

$$h(\mu_{ij}) = \underset{\sim}{x}_{ij}^{T}\underset{\sim}{\beta} \qquad \text{and} \qquad var(y_{ij}|\underset{\sim}{x}_{ij}) = g(\mu_{ij})\cdot\phi, \tag{3.1}$$

where i, $i = 1,\ldots,I$, denotes the cluster index; j, $j = 1,\ldots,n_i$, denotes the observation index; y_{ij} is the response variable; $\underset{(p\,x\,1)}{\underset{\sim}{x}_{ij}}$ is the fixed covariate matrix, $E(y_{ij}|\underset{\sim}{x}_{ij}) = \mu_{ij}$ is the marginal expectation; $\underset{(p\,x\,1)}{\underset{\sim}{\beta}}$ is the marginal parameter vector, $h(\cdot)$ is a monotone and differentiable link function, $g(\cdot)$ is the variance function, and ϕ is a scale parameter (Zeger, Liang and Albert, 1988).

The exponential family of distributions, in general, can be written as

$$f(y,\theta,\phi) = c(y,\phi)exp[\{y\theta - b(\theta)\}/a(\phi)] \tag{3.2}$$

where the respective mean and variance of a response, y, are denoted by

$$E(Y) = \mu = \mu(\theta) = b'(\theta) \quad Var(Y) = b''(\theta)\cdot a(\phi) = a(\phi)\cdot v\{\mu(\theta)\}. \tag{3.3}$$

The connection between the marginal model and the exponential family notation is the following, where for now we repress the j index presuming all clusters are of size $n_i = 1$:

$$\theta_i = (h\circ\mu)^{-1}(\underset{\sim}{x}_i^T\underset{\sim}{\beta}) \text{ and } \quad v\{\mu(\theta)\} = g_i\{\mu_i\}. \tag{3.4}$$

To simplify the notation we denote $(h\circ\mu)^{-1}(\cdot)$ by $\kappa(\cdot)$.

We begin by developing the score vector and information matrix for the independence case, $n_i = 1$. The β-dependent part of the marginal model log-likelihood can be written as

$$ln\boldsymbol{L_I}(\underset{\sim}{\boldsymbol{\beta}}) = \sum_{i=1}^{I}[Y_i\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta}) - b\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}]. \tag{3.5}$$

The score vector is expressed as

$$\boldsymbol{U_I}(\underset{\sim}{\boldsymbol{\beta}}) = (\partial/\partial\underset{\sim}{\boldsymbol{\beta}})ln\boldsymbol{L_I}(\underset{\sim}{\boldsymbol{\beta}}) = \sum_{i=1}^{I}\frac{\{Y_i-\mu_i(\underset{\sim}{\beta})\}}{[h'\{\mu_i(\underset{\sim}{\beta})\}g_i(\underset{\sim}{\beta})]}\underset{\sim}{x}_i = \underset{\sim}{\boldsymbol{X}}_{\boldsymbol{I}}^{\boldsymbol{T}}\boldsymbol{D_I^{-1}}(\underset{\sim}{\boldsymbol{\beta}})\boldsymbol{S_I}(\underset{\sim}{\boldsymbol{\beta}}) \tag{3.6}$$

where we define $\underset{\sim}{\widehat{\beta}}_I$ as the solution to (3.6), $\underset{\sim}{\boldsymbol{X}}_{\boldsymbol{I}}^{\boldsymbol{T}} = [\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_I]$,

$\boldsymbol{D_I}(\underset{\sim}{\boldsymbol{\beta}}) = diag[[h'\{\mu_1(\underset{\sim}{\beta})\}g_1(\underset{\sim}{\beta})],\ldots,[h'\{\mu_I(\underset{\sim}{\beta})\}g_I(\underset{\sim}{\beta})]]$, and

$\boldsymbol{S_I}(\underset{\sim}{\beta}) = [\{Y_1 - \mu_1(\underset{\sim}{\beta})\}, \ldots, \{Y_I - \mu_I(\underset{\sim}{\beta})\}]^T$. The second derivative of the log-likelihood is the information matrix defined as

$$\boldsymbol{U_I'}(\underset{\sim}{\beta}) = -(\partial^2/\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}^t)ln\boldsymbol{L_I}(\underset{\sim}{\beta}) = \sum_{i=1}^{I}[h'\{\mu_i(\underset{\sim}{\beta})\}]^{-2}\{g_i(\underset{\sim}{\beta})\}^{-1}\underset{\sim}{x}_i\underset{\sim}{x}_i^T + \sum_{i=1}^{I}\{Y_i - \mu_i(\underset{\sim}{\beta})\}\left(\frac{h''\{\mu_i(\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2} + \frac{b^{(3)}\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2\{g_i(\underset{\sim}{\beta})\}^3}\right)\underset{\sim}{x}_i\underset{\sim}{x}_i^T. \quad (3.7)$$

The expected information matrix is denoted as

$$\boldsymbol{J_I}(\underset{\sim}{\beta}) = E\{-(\partial^2/\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}^T)ln\boldsymbol{L_I}(\underset{\sim}{\beta})\} = \sum_{i=1}^{I}[h'\{\mu_i(\underset{\sim}{\beta})\}]^{-2}\{g_i(\underset{\sim}{\beta})\}^{-1}\underset{\sim}{x}_i\underset{\sim}{x}_i^T, \quad (3.8)$$

which is conditional upon the covariates. Therefore, the second derivative of the log-likelihood can be re-expressed as a sum of two pieces, the expected information and a remainder term

$$\boldsymbol{U_I'}(\underset{\sim}{\beta}) = -(\partial^2/\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}^t)ln\boldsymbol{L_I}(\underset{\sim}{\beta}) = E[-(\partial^2/\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}^t)ln\boldsymbol{L_I}(\underset{\sim}{\beta})] + \sum_{i=1}^{I}\{Y_i - \mu_i(\underset{\sim}{\beta})\}\left(\frac{h''\{\mu_i(\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2} + \frac{b^{(3)}\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2\{g_i(\underset{\sim}{\beta})\}^3}\right)\underset{\sim}{x}_i\underset{\sim}{x}_i^T$$

$$= \boldsymbol{J_I}(\underset{\sim}{\beta}) + \boldsymbol{R_I}(\underset{\sim}{\beta}), \quad (3.9)$$

where $\boldsymbol{R_I}(\underset{\sim}{\beta}) = \sum_{i=1}^{I}\{Y_i - \mu_i(\underset{\sim}{\beta})\}\left(\frac{h''\{\mu_i(\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2} + \frac{b^{(3)}\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2\{g_i(\underset{\sim}{\beta})\}^3}\right)\underset{\sim}{x}_i\underset{\sim}{x}_i^T$. Weak consistency of (3.9) is required in the proof of the asymptotic normality of $\underset{\sim}{\widehat{\beta}}_I$. To show weak consistency, first we need to assume that

$$\lim_{I\to\infty}\ \frac{1}{I}\boldsymbol{J_I}(\underset{\sim}{\beta}) = \boldsymbol{J}(\underset{\sim}{\beta}) \text{ is a finite and positive definite matrix,} \quad (3.10)$$

(Sen and Singer, 1993). This essentially states that the expected information matrix converges to some constant matrix and the dependence on the covariates vanishes as $I \to \infty$. Note that the matrix of remainder terms, $\boldsymbol{R_I}(\underset{\sim}{\beta})$, is written as a sum of independent, centered random variables that have finite variances $g_i\{\mu_i(\underset{\sim}{\beta})\}$ and non-stochastic matrix coefficients defined as

$$\underset{\sim}{G}_i = \left(\frac{h''\{\mu_i(\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2} + \frac{b^{(3)}\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}}{(h'\{\mu_i(\underset{\sim}{\beta})\})^2\{g_i(\underset{\sim}{\beta})\}^3} \right) \underset{\sim}{x}_i \underset{\sim}{x}_i^t, \tag{3.11}$$

which are conditional upon the x's (Sen and Singer, 1993).

Lemma 1: If the variance of the (l,m) element of the (pxp) matrix, $I^{-1} \cdot \boldsymbol{R_I}(\underset{\sim}{\beta})$, goes to zero then the variance of every element of the matrix will be zero

$$Var(I^{-1} \cdot R_{I(l,m)}(\underset{\sim}{\beta})) = \lim_{I\to\infty} \; I^{-2} \sum_{i=1}^{I} g_{i(l,m)}\{\mu_i(\underset{\sim}{\beta})\}(G^2_{i(l,m)}) \to 0.$$

Proof:

$$Var(I^{-1} \cdot R_{I(l,m)}(\underset{\sim}{\beta})) = E(I^{-1} \cdot R_{I(l,m)}(\underset{\sim}{\beta}))^2$$

$$= I^{-2} \cdot E(\sum_{i=1}^{I}\{Y_i - \mu_i(\underset{\sim}{\beta})\}G_{i(l,m)})^2$$

$$\leq I^{-2} \cdot E(\sum_{i=1}^{I}\{Y_i - \mu_i(\underset{\sim}{\beta})\})^2(G^2_{i(l,m)})$$

$$= I^{-2} \cdot (\sum_{i=1}^{I} E(r_i^2)(G^2_{i(l,m)}))$$

$$\doteq I^{-1} \cdot g_i\{\mu_i(\underset{\sim}{\beta})\}(G^2_{i(l,m)}) = O(I^{-1}) \text{ as } I \to \infty.$$

The Markov Weak Law of Large Numbers (Theorem 2.3.7, Sen and Singer, 1993) is satisfied by showing that $\boldsymbol{R_I}(\underset{\sim}{\boldsymbol{\beta}})$ goes to zero in the second mean ($\delta = 1$).

Markov Weak Law of Large Numbers:

Let X_k, $k \geq 1$, be independent random variables, such that $\mu_k = E(X_k)$, $k \geq 1$ exist, and for some δ such that $0 < \delta \leq 1$, along with the existence of $E|X_k - \mu_k|^{1+\delta}$, suppose that

$$n^{-1-\delta}\sum_{k=1}^{n} E|X_k - \mu_k|^{1+\delta} = \rho_n(\delta) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Then $\overline{X}_n - E(\overline{X}_n) \rightarrow 0$.

The second mean condition was easy to show, however, it is not necessary and the theorem can be satisfied with a less stringent condition, such as $\delta < 1$. Applying Chebyshev's Inequality to $\boldsymbol{R_I}(\underset{\sim}{\boldsymbol{\beta}})$ gives the following, $I^{-1} \cdot \boldsymbol{R_I}(\underset{\sim}{\boldsymbol{\beta}}) \xrightarrow{p} \underset{\sim}{0}$, which in turn provides the necessary piece to establish the required weak consistency of the second derivative of the log-likelihood,

$$I^{-1}\{-(\partial^2/\partial\underset{\sim}{\beta}\,\partial\underset{\sim}{\beta}^{t})ln\boldsymbol{L_I}(\underset{\sim}{\boldsymbol{\beta}})\} = I^{-1}(\boldsymbol{J_I}(\underset{\sim}{\boldsymbol{\beta}}) + \boldsymbol{R_I}(\underset{\sim}{\boldsymbol{\beta}}))$$
$$\doteq I^{-1} \cdot \boldsymbol{J}(\underset{\sim}{\boldsymbol{\beta}}) + \boldsymbol{o(1)} \text{ in the } 2^{\text{nd}} \text{ mean} \qquad (3.12)$$

as $I \rightarrow \infty$. These assumptions ensure that the required consistency holds when the model is true (Sen and Singer, 1993).

Furthermore, the expected information matrix, $\boldsymbol{J_I}(\underset{\sim}{\boldsymbol{\beta}})$, depends on $h(\cdot)$, $g(\cdot)$, and the explanatory variables, $\underset{\sim}{x}_i$. Additional regularity and compactness conditions are required to establish the asymptotic properties of the estimators. Also a compactness

condition on the second derivative of the log-likelihood function is required. Define the following

$$B(\delta) = \{ \underset{\sim}{\beta}^* \in \mathbb{R}^p : \| \underset{\sim}{\beta}^* - \underset{\sim}{\beta} \| < \delta \}, \ \delta \downarrow 0, \tag{3.13}$$

$$\boldsymbol{w_{1i}}(\underset{\sim}{\beta}) = [h'\{\mu_i(\underset{\sim}{\beta})\}]^{-2}\{g_i(\underset{\sim}{\beta})\}^{-1}, \text{ and} \tag{3.14}$$

$$\begin{aligned} \boldsymbol{w_{2i}}(\underset{\sim}{\beta}) = \mu_i(\underset{\sim}{\beta})\{h''\{\mu_i(\underset{\sim}{\beta})\}[h'\{\mu_i(\underset{\sim}{\beta})\}]^{-2} + \\ b^{(3)}\{\kappa(\underset{\sim}{x}_i^T\underset{\sim}{\beta})\}[h'\{\mu_i(\underset{\sim}{\beta})\}]^{-2}\{g_i(\underset{\sim}{\beta})\}^{-3}\}. \end{aligned} \tag{3.15}$$

Excluding the covariates, the $\boldsymbol{w_{1i}}(\underset{\sim}{\beta})$ is the coefficient of the expected information matrix, $\boldsymbol{J_I}(\underset{\sim}{\beta})$, and the $\boldsymbol{w_{2i}}(\underset{\sim}{\beta})$ is the coefficient of the remainder term of the information matrix, $\boldsymbol{R_I}(\underset{\sim}{\beta})$. Then the required compactness conditions may be expressed as

i) for $k = 1, 2$, as $\delta \downarrow 0$

$$\sup_{\underset{\sim}{\beta}^* \in B(\delta)} I^{-1}\sum_{i=1}^{I} \| \{\boldsymbol{w_{ki}}(\underset{\sim}{\beta}^*) - \boldsymbol{w_{ki}}(\underset{\sim}{\beta})\}\underset{\sim}{x}_i\underset{\sim}{x}_i^T \| \to 0$$

Since we are taking the *sup* over $B(\delta)$, then i) holds for the (l, m) element of $\boldsymbol{w_{ki}}(\cdot)$ as given below

$$I^{-1}\sum_{i=1}^{I} \| \{w_{ki(l,m)}(\underset{\sim}{\beta}^*) - w_{ki(l,m)}(\underset{\sim}{\beta})\}x_{i(l,m)}x_{i(l,m)}^T \| \to 0$$

ii) as $\delta \downarrow 0$

$$E_{\underset{\sim}{\beta}}\{\sup_{\underset{\sim}{\beta}^* \in B(\delta)} I^{-1}\sum_{i=1}^{I}|Y_i|\|\boldsymbol{w_{2i}}(\underset{\sim}{\beta}^*) - \boldsymbol{w_{2i}}(\underset{\sim}{\beta})\underset{\sim}{x}_i\underset{\sim}{x}_i^T\|\} \rightarrow 0$$

Since we are taking the *sup* over $B(\delta)$, then (ii) holds for the (l, m) element of $\boldsymbol{w_{2i}}(\cdot)$ as given below

$$E_{\underset{\sim}{\beta}}\{I^{-1}\sum_{i=1}^{I}|Y_i|\|w_{2i(l,m)}(\underset{\sim}{\beta}^*) - w_{2i(l,m)}(\underset{\sim}{\beta})x_{i(l,m)}x_{i(l,m)}^T\|\} \rightarrow 0$$

(These conditions can be found in Sen and Singer, 1993.) The second compactness condition (ii) is necessary for proving Proposition 1 as stated below.

A standard proof of the asymptotic normality of the MLE estimator under a GLM [Proposition 1 below] is found in most theoretical textbooks (Cox and Hinkley, 1974; Sen and Singer, 1993). We include the proof because features of it will be reused in establishing the asymptotic normality of the standardized WCR estimator.

Proposition 1: Under a GLM as described above if $\underset{\sim}{\hat{\beta}}_I$ is the MLE with sample size I, then $\sqrt{I}(\underset{\sim}{\hat{\beta}}_I - \underset{\sim}{\beta}) \sim N(\underset{\sim}{0}, \boldsymbol{J}^{-1}(\underset{\sim}{\beta}))$ as $I \rightarrow \infty$.

Proof:

We begin by taking the Taylor series expansion of the log-likelihood around the true parameter $\underset{\sim}{\beta}$. Let $\|\underset{\sim}{u}\| < K$, where $\underset{\sim}{u} \in \mathbb{R}^p$ such that $\|\underset{\sim}{u}\|$ is less than K for positive K. Define the following function of $\underset{\sim}{u}$

$$\lambda_I(\underset{\sim}{u}) = lnL_I(\underset{\sim}{\beta} + I^{-1/2}\underset{\sim}{u}) - lnL_I(\underset{\sim}{\beta})$$

$$= \frac{1}{\sqrt{I}}\underset{\sim}{u}^T U_I(\underset{\sim}{\beta}) + \frac{1}{2I}\underset{\sim}{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}^*}\underset{\sim}{u} \tag{3.16}$$

where $\underset{\sim}{\beta}^*$ belongs to the line segment joining $\underset{\sim}{\beta}$ and $\underset{\sim}{\beta} + I^{-1/2}\underset{\sim}{u}$ (Sen and Singer, 1993). Let

$$Z_I(\underset{\sim}{u}) = \frac{1}{2I}\{\underset{\sim}{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}^*}\underset{\sim}{u} - \underset{\sim}{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}}\underset{\sim}{u} +$$
$$\underset{\sim}{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}}\underset{\sim}{u} + \underset{\sim}{u}^T J_I(\underset{\sim}{\beta})\underset{\sim}{u}\} \tag{3.17}$$

be the remainder term where $J_I(\underset{\sim}{\beta})$ is defined as in (3.8). Therefore, we may rewrite (3.16) as

$$\lambda_I(\underset{\sim}{u}) = \frac{1}{\sqrt{I}}\underset{\sim}{u}^T U_I(\underset{\sim}{\beta}) - \frac{1}{2I}\underset{\sim}{u}^T J_I(\underset{\sim}{\beta})\underset{\sim}{u} + Z_I(\underset{\sim}{u}). \tag{3.18}$$

Notice that

$$\sup_{\|\underset{\sim}{u}\|<K} \|Z_I(\underset{\sim}{u})\| \le \frac{1}{2}\sup_{\underset{\sim}{\beta}^* \in B(K/\sqrt{I})} \left\| \frac{1}{I}\frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}^*} - \frac{1}{I}\frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta})\Big|_{\underset{\sim}{\beta}} \right\|$$
$$+ \frac{K^2}{2}\left\| \frac{1}{I}\frac{\partial^2}{\partial\beta\partial\beta^T} lnL_I(\underset{\sim}{\beta}) + \frac{1}{I}J_I(\underset{\sim}{\beta}) \right\| \tag{3.19}$$

can be shown to be $o(1)$ in the 1st mean uniformly in $\underset{\sim}{u}$, $\|\underset{\sim}{u}\| < K$ by conditions (i) and (ii). Now we maximize $\lambda_I(\underset{\sim}{u})$ with respect to $\underset{\sim}{u}$ by using (3.10) ($\frac{1}{I}J_I(\underset{\sim}{\beta}) = J(\underset{\sim}{\beta})$ is a finite and positive definite matrix as $I \to \infty$) to obtain

$$\hat{\underset{\sim}{u}} = \frac{1}{\sqrt{I}} \boldsymbol{J}^{-1}(\underset{\sim}{\boldsymbol{\beta}}) \boldsymbol{U}_I(\underset{\sim}{\boldsymbol{\beta}}) + \boldsymbol{o}(1) \text{ in the } 1^{st} \text{ mean.} \tag{3.20}$$

The terms summed on the right-hand side of (3.20) are independent and identically distributed random variables. Therefore, the central limit theorem can be applied to show that for large I

$$\frac{1}{\sqrt{I}} \boldsymbol{J}^{-1}(\underset{\sim}{\boldsymbol{\beta}}) \boldsymbol{U}_I(\underset{\sim}{\boldsymbol{\beta}}) \sim \boldsymbol{N}_p(\underset{\sim}{0}, \boldsymbol{J}^{-1}(\underset{\sim}{\boldsymbol{\beta}})). \tag{3.21}$$

Looking at the definition of $\boldsymbol{\lambda}_I(\underset{\sim}{\boldsymbol{u}})$, we see that $\hat{\underset{\sim}{u}} \doteq \sqrt{I}(\hat{\underset{\sim}{\beta}}_I - \underset{\sim}{\beta})$ (Sen and Singer, 1993), consequently,

$$\sqrt{I}(\hat{\underset{\sim}{\beta}}_I - \underset{\sim}{\beta}) = \frac{1}{\sqrt{I}} \boldsymbol{J}^{-1}(\underset{\sim}{\boldsymbol{\beta}}) \boldsymbol{U}_I(\underset{\sim}{\boldsymbol{\beta}}) + \boldsymbol{o}(1) \text{ in the } 1^{st} \text{ mean.} \tag{3.22}$$

Using Slutsky's Theorem (Theorem 3.4.2, Sen and Singer, 1993), the desired result, Proposition 1, is obtained

$$\sqrt{I}(\hat{\underset{\sim}{\beta}}_I - \underset{\sim}{\beta}) \sim \boldsymbol{N}_p(\underset{\sim}{0}, \boldsymbol{J}^{-1}(\underset{\sim}{\boldsymbol{\beta}})) \text{ as } I \to \infty \tag{3.23}$$

3.3 WCR Sampling Scheme

For clarity, the steps of the WCR resampling algorithm are explicitly stated, as well as displayed in Figure 3.1. The WCR method randomly samples one observation and its corresponding covariate vector from each cluster with replacement. The resampled data set is analyzed by a generalized linear model, e.g. logistic model, since the observations within a resampled data set are independent. This process is repeated a large number of times, say Q, $q = 1, \ldots, Q$, where each of the Q analyses provides a consistent estimate of the marginal parameter of interest. Conditional on the data, the

estimates from the Q resample-based analyses, $\widehat{\underset{\sim}{\beta}}(R;q)$, $q = 1, \ldots, Q$, are independent and identically distributed. However, unconditionally the Q resampled data sets are dependent, since the resampled-data sets contain correlated and overlapping observations. The overall estimate of the effect of interest is found by computing the average of the Q resample-based estimates. The dependency does not affect the WCR estimator, but it is implicitly taken into account by the variance formula. The following display, Figure 3.1, provides visual representation of the WCR sampling scheme.

Figure 3.1. WCR Sampling Scheme

Consider a hypothetical data set consisting of I clusters ...

$$\underset{\sim}{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \end{bmatrix} \quad \underset{\sim}{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} \quad \underset{\sim}{y}_3 = [y_{31}] \quad \underset{\sim}{y}_4 = \begin{bmatrix} y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{bmatrix} \quad \cdots \quad \underset{\sim}{y}_I = \begin{bmatrix} y_{I1} \\ y_{I2} \end{bmatrix}$$

By repeated (Q times), random sampling of one observation from each of the above I clusters, a series of Q, $q = 1, \ldots, Q$, resampled data sets $R(q)$ of size I and resample-based estimates ($\underset{\sim}{\widehat{\beta}}\,(R;q)$) are obtained ...

$$R(1) = \begin{vmatrix} y_{11}, \underset{\sim}{x}_{11} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{44}, \underset{\sim}{x}_{44} \\ \vdots \\ y_{I1}, \underset{\sim}{x}_{I1} \end{vmatrix} \underset{\rightarrow}{\text{GLM}}\ \underset{\sim}{\widehat{\beta}}\,(R;1) \quad \cdots \quad R(Q) = \begin{vmatrix} y_{12}, \underset{\sim}{x}_{12} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{42}, \underset{\sim}{x}_{42} \\ \vdots \\ y_{I2}, \underset{\sim}{x}_{I2} \end{vmatrix} \underset{\rightarrow}{\text{GLM}}\ \underset{\sim}{\widehat{\beta}}\,(R;Q)$$

The WCR estimator ($\underset{\sim}{\overline{\beta}}$) is obtained by averaging the resample-based estimates across the resamples...

$$\underset{\sim}{\overline{\beta}} = \frac{\sum_{q=1}^{Q} \underset{\sim}{\widehat{\beta}}(R;q)}{Q}.$$

Unlike the GEE parameter, the WCR marginal parameter is cluster-based since the cluster is the unit of analysis. The cluster-based parameter results from the sampling scheme of WCR, which gives equal weight to all clusters in an analysis. In contrast,

traditional marginal model parameters are observation-based since the unit of analysis is the observation and the weight of the cluster increases with its size.

3.4. Asymptotic Normality of a Standardized WCR Estimator

The estimator from a resampled data set, $\widehat{\beta}(R)$, first is shown to be consistent for the true WCR parameter, β.

Proposition 2: Define $E_i[\,\cdot\,]$ as the expectation across clusters and $E_{j|i,X_{ij}=x}(\,\cdot\,)$ as the expectation across observations within a cluster for observations with $X_{ij} = x$, then $\widehat{\beta}(R)$ is consistent for β where

$$\beta = h\{E_i[E_{j|i,X_{ij}=x+1}(Y_{ij}|X_{ij} = x+1)]\} - h\{E_i[E_{j|i,X_{ij}=x}(Y_{ij}|X_{ij} = x)]\} \qquad (3.24)$$

The above definition in (3.24) represents the sampling that occurs with WCR. The estimator, $\widehat{\beta}(R)$, is the standard maximum likelihood estimator from a generalized linear model given the sampling scheme of WCR. The consistency of $\widehat{\beta}(R)$ follows from the usual asymptotics of generalized linear models (Fahrmeir and Kaufmann, 1985).

We will now establish the asymptotic normality of the standardized version of the WCR estimator (Proposition 3). We will make use of the material introduced in Section 3.2 with the addition of an index or superscript of 'q' to denote the q^{th} resampled data set. For example, we denote $\boldsymbol{J_I(\underset{\sim}{\beta}\,;q)}$ as the expected information matrix for a sample size of I from the q^{th} resampled data set or $\boldsymbol{Z_I^{(q)}(\underset{\sim}{u}\,)}$ as the remainder term defined in (3.17) for the q^{th} resampled data set.

Proposition 3: Let the WCR estimator be defined as $\overline{\underset{\sim}{\beta}} = \frac{1}{Q}\sum_{q=1}^{Q}\widehat{\underset{\sim}{\beta}}(R;q)$ for Q very large. As the number of clusters, I, goes to ∞, $\sqrt{I}(\overline{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \sim N(\underset{\sim}{0}, \underset{\sim}{\Sigma}), \underset{\sim}{\Sigma}$ unspecified, but finite and positive-definite.

Remark:

The estimators of the parameter vector of interest for the Q resamples, $\widehat{\underset{\sim}{\beta}}(R;q)$, $q = 1, \ldots, Q$, are correlated. The resampled-data sets are not independent since the observations within a cluster are assumed to be correlated and there is overlap in the resampled data sets. Therefore, even though the observations within a specific resampled-data set are independent because they are from distinct clusters and each $\widehat{\underset{\sim}{\beta}}(R)$ is asymptotically normal by standard theory for GLM's (Proposition 1), the WCR estimator is the average of dependent MLE's. Therefore, the Central Limit Theorem is not directly applicable to the WCR estimator.

Proof:

To establish asymptotic normality, the idea is to rewrite the average of the Q resample-based score statistics as the sum of independent pieces so that a Central Limit Theorem may be applied. By using parts of the proof of Proposition 1, the asymptotic normality of a standardized version of the WCR estimator can be established and subsequently a consistent variance estimator can be defined.

We note that the proof of the asymptotic normality in Proposition 1 holds for each $\widehat{\underset{\sim}{\beta}}(R;q)$, $q = 1, \ldots, Q$. Therefore, we define a function similar to (3.19) for each $q, q = 1, \ldots, Q$

$$\lambda_I^{(q)}(\underset{\sim}{u}) = \frac{1}{\sqrt{I}}\underset{\sim}{u}^T \boldsymbol{U}_I(\underset{\sim}{\beta};q) - \frac{1}{2I}\underset{\sim}{u}^T \boldsymbol{J}_I(\underset{\sim}{\beta};q)\underset{\sim}{u} + \boldsymbol{Z}_I^{(q)}(\underset{\sim}{u}) \qquad (3.25)$$

where each $Z_I^{(q)}(\underset{\sim}{u})$ is $o_p(1)$ by (3.19). Averaging (3.25) across $q, q = 1, \ldots, Q$,

$$\frac{1}{Q}\sum_{q=1}^{Q}\lambda_I^{(q)}(\underset{\sim}{u}) = \frac{1}{Q}\sum_{q=1}^{Q}\{\frac{1}{\sqrt{I}}\underset{\sim}{u}^T U_I(\underset{\sim}{\beta};q) - \frac{1}{2I}\underset{\sim}{u}^T J_I(\underset{\sim}{\beta};q)\underset{\sim}{u} + Z_I^{(q)}(\underset{\sim}{u})\} \quad (3.26)$$

and maximizing (3.26) with respect to $\underset{\sim}{u}$, as in Proposition 1, we get

$$\hat{\underset{\sim}{u}} = \frac{\sqrt{I}}{Q}\sum_{q=1}^{Q}\{J_I^{-1}(\underset{\sim}{\beta};q)\cdot U_I(\underset{\sim}{\beta};q) + o(1) \text{ in the } 1^{st} \text{ mean}\}$$

$$= \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}}\sum_{q=1}^{Q}U_I(\underset{\sim}{\beta};q) + o(1) \text{ in the } 1^{st} \text{ mean}, \quad (3.27)$$

since $\frac{1}{Q}\sum_{q=1}^{Q}o(1)$ in the 1^{st} mean $= o(1)$ in the 1^{st} mean. For each q, the consistency shown in (3.12) holds,

$$I^{-1}\cdot(-U_I'(\underset{\sim}{\beta};q)) = I^{-1}\cdot J_I(\underset{\sim}{\beta};q) + I^{-1}\cdot R_I(\underset{\sim}{\beta};q) = J(\underset{\sim}{\beta}) + o(1) \text{ in the } 2^{nd} \text{ mean},$$

because each comes from the same joint distribution. We are able to take the derivative inside the summation in (3.26) because of the uniform continuity established by the regularity conditions (i)-(ii) previously stated (see Theorem 2.4.3 of Casella and Berger, 1990 for more details).

To average over $q, q = 1, \ldots, Q$, it is not sufficient to only have each $Z_I^{(q)}(\underset{\sim}{u}) = o(1)$ in the 1^{st} mean since the variance of an average of dependent terms produces covariances which may or may not be negligible. To eliminate the possibility of the remainder terms, the $Z_I^{(q)}(\underset{\sim}{u})$'s, contributing to the variance, it is necessary to use the stated regularity conditions. We begin by showing that the variance of the average of the (l, m) elements of the $Z_I^{(q)}(\underset{\sim}{u})$'s for $q = 1, \ldots, Q$ goes to zero

$$Var(\frac{1}{Q}\sum_{q=1}^{Q} Z^{(q)}_{I(l,m)}(\underset{\sim}{u})) \leq E(\frac{1}{Q}\sum_{q=1}^{Q} Z^{(q)}_{I(l,m)}(\underset{\sim}{u}))^2$$

$$= Q^{-2} \cdot \sum_{q=1}^{Q} E(Z^{(q)}_{I(l,m)}(\underset{\sim}{u}))^2 + Q^{-2}\sum_{q\neq q'}^{Q} E(Z^{(q)}_{I(l,m)}(\underset{\sim}{u}), Z^{(q')}_{I(l,m)}(\underset{\sim}{u}))$$

$$= O(Q^{-1}) + O(I^{-1}) = o(1) \tag{3.28}$$

using the Cauchy-Schwarz Inequality and conditions (i)-(ii).

We can rewrite the average on the right-hand side of (3.28) (this is the average of Q score vectors) as the sum of independent and identically distributed random variables

$$\overline{U}(\underset{\sim}{\beta}) = \frac{1}{Q}\sum_{q=1}^{Q} U_I(\underset{\sim}{\beta}; q)$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\sum_{i=1}^{I} U_i(\underset{\sim}{\beta}; q)$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\{\sum_{i=1}^{I} \frac{\{Y_{iq}-\mu_{iq}(\underset{\sim}{\beta})\}}{[h'\{\mu_{iq}(\underset{\sim}{\beta})\}g_{iq}(\underset{\sim}{\beta})]}\underset{\sim}{x}_{iq}\}$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\{\sum_{i=1}^{I}\underset{\sim}{x}^{T}_{iq} d^{-1}_{iq} r_{iq}\}$$

$$= \frac{1}{Q}\sum_{q=1}^{Q}\{\underset{\sim}{x}^{T}_{1q} d^{-1}_{1q} r_{1q} + \ldots + \underset{\sim}{x}^{T}_{Iq} d^{-1}_{Iq} r_{Iq}\}$$

$$= \frac{1}{Q}\{(\underset{\sim}{x}^{T}_{11} d^{-1}_{11} r_{11} + \ldots + \underset{\sim}{x}^{T}_{I1} d^{-1}_{I1} r_{I1}) + \ldots + (\underset{\sim}{x}^{T}_{1Q} d^{-1}_{1Q} r_{1Q} + \ldots + \underset{\sim}{x}^{T}_{IQ} d^{-1}_{IQ} r_{IQ})\}$$

$$= \frac{1}{Q}\{(\underset{\sim}{x}_{11}^{T} d_{11}^{-1} r_{11} + \ldots + \underset{\sim}{x}_{1Q}^{T} d_{1Q}^{-1} r_{1Q}) + \ldots +$$
$$(\underset{\sim}{x}_{I1}^{T} d_{I1}^{-1} r_{I1} + \ldots + \underset{\sim}{x}_{IQ}^{T} d_{IQ}^{-1} r_{IQ})\}$$

$$= \frac{1}{Q}\{\underset{\sim}{X}_{I1}^{*T} D^{*-1}_{I1}(\underset{\sim}{\beta}) S^{*}_{I1}(\underset{\sim}{\beta}) + \ldots + \underset{\sim}{X}_{II}^{*T} D^{*-1}_{II}(\underset{\sim}{\beta}) S^{*}_{II}(\underset{\sim}{\beta})\}$$

$$= \frac{1}{Q}\sum_{i=1}^{I} \underset{\sim}{X}_{Ii}^{*T} D^{*-1}_{Ii}(\underset{\sim}{\beta}) S^{*}_{Ii}(\underset{\sim}{\beta}). \tag{3.29}$$

where $d_{iq} = [h'\{\mu_{iq}(\underset{\sim}{\beta})\} g_{iq}(\underset{\sim}{\beta})]$, $r_{iq} = \{Y_{iq} - \mu_{iq}(\underset{\sim}{\beta})\}$, $\underset{\sim}{X}_{Ii}^{*T} = [\underset{\sim}{x}_{i1}, \ldots, \underset{\sim}{x}_{iQ}]$, $D^{*}_{Ii}(\underset{\sim}{\beta}) = diag[d_{i1}, \ldots, d_{iQ}]$, and $S^{*}_{Ii}(\underset{\sim}{\beta}) = [r_{i1}, \ldots, r_{iQ}]^{T}$. Therefore, since (3.29) is the sum of independent and identically random variables, we may appeal to a Central Limit Theorem (Sen and Singer, 1993) to establish the asymptotic normality of the standardized average of the score vectors across the resamples.

From Proposition 1 and (3.27), we can write

$$\overline{\underset{\sim}{\beta}} = \frac{1}{Q}\sum_{q=1}^{Q} \underset{\sim}{\hat{\beta}}_{I}(R; q)$$

$$= \underset{\sim}{\beta} + I^{-1/2}\{\frac{J^{-1}(\beta)}{\sqrt{IQ}}\sum_{q=1}^{Q}(U_I(\underset{\sim}{\beta}; q) + \frac{o_p(I^{1/2})}{\sqrt{I}})\}.$$

Therefore, we have by (3.28)

$$\sqrt{I}(\overline{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) = \frac{J^{-1}(\beta)}{\sqrt{IQ}}\sum_{q=1}^{Q} U_I(\underset{\sim}{\beta}; q) + o_p(1). \tag{3.30}$$

Using Slutsky's Theorem, the desired result follows as $I \rightarrow \infty$,

$$\sqrt{I}(\underset{\sim}{\hat{\beta}} - \underset{\sim}{\beta}) \sim \boldsymbol{N_p}(\underset{\sim}{0}, \underset{\approx}{\Sigma}), \tag{3.31}$$

where $\underset{\approx}{\Sigma}$ is an unspecified, but finite and positive-definite matrix as $I \rightarrow \infty$, provided we can verify that $\underset{\approx}{\Sigma} = Var(\frac{1}{Q\sqrt{I}}\sum_{q}^{Q}\sum_{i}^{I}\boldsymbol{U_i}(\underset{\sim}{\beta}; \boldsymbol{q}))$ goes to a non-zero finite limit. We can rewrite

$$Var(\frac{1}{Q\sqrt{I}}\sum_{q}^{Q}\sum_{i}^{I}\boldsymbol{U_i}(\underset{\sim}{\beta}; \boldsymbol{q})) = \frac{1}{Q^2 I}[Var\sum_{q}^{Q}\boldsymbol{U_I}(\underset{\sim}{\beta}; \boldsymbol{q})]$$

$$= \frac{1}{Q^2 I}[\sum_{q}^{Q} Var(\boldsymbol{U_I}(\underset{\sim}{\beta}; \boldsymbol{q})) + \sum_{q \neq q'} Cov(\boldsymbol{U_I}(\underset{\sim}{\beta}; \boldsymbol{q}), \boldsymbol{U_I}(\underset{\sim}{\beta}; \boldsymbol{q'}))]. \tag{3.32}$$

To simplify notation without loss of generality, assume that $n_i = n \; \forall i$ and $Q \rightarrow \infty$ for each I. Under random sampling, each of the n^I possible outcomes is equally likely. Under this scenario, we group the pairs of score vectors into those with 'l' terms in common. The following table, Table 3.1, displays the number of possible matches of 'l' terms in common in pairs of score vectors.

Table 3.1. All Possible Pairs of Score Vectors by the Number of Common Terms

# common terms	# of possible matches
0	$n^I(n-1)^I$
1	$I \cdot n^I(n-1)^{I-1}$
l	$\binom{I}{l} n^I(n-1)^{I-l}$
$\vdots$	$\vdots$
$I-1$	$I \cdot n^I(n-1)$
I	n^I

Denoting the covariance between a pair of score vectors with 'l' terms in common by $\underset{\approx}{\Delta}_l$, we can rewrite (3.32) as

$$\frac{n^I\underset{\approx}{\Delta}_I+I\cdot n^I(n-1)\underset{\approx}{\Delta}_{I-1}+\ldots+\binom{I}{i}n^I(n-1)^{I-i}\underset{\approx}{\Delta}_i+\ldots+I\cdot n^I(n-1)^{I-1}\underset{\approx}{\Delta}_1+n^I(n-1)^I\underset{\approx}{\Delta}_0}{I\cdot n^{2I}}. \tag{3.33}$$

We know that the covariance between pairs of score vectors with no terms in common is less than or equal to that for pairs with one term in common and so on. Consequently the covariance matrices can be ranked as follows,

$$\underset{\approx}{\Delta}_0 \le \underset{\approx}{\Delta}_1 \le \cdots \le \underset{\approx}{\Delta}_i \le \cdots \le \underset{\approx}{\Delta}_{I-1} \le \underset{\approx}{\Delta}_I \tag{3.34}$$

where we assume that $\underset{\approx}{\Delta}_0 > 0$ and $\underset{\approx}{\Delta}_I$ is by definition the variance of $\boldsymbol{U}_I(\underset{\sim}{\boldsymbol{\beta}})$. The sum of covariance matrices in (3.33) can be written as the following binomial sum

$$\frac{1}{I}\sum_{i=0}^{I}\binom{I}{i}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{I-i}\underset{\approx}{\Delta}_i \ge \frac{1}{I}\sum_{i=0}^{I}\binom{I}{i}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{I-i}\underset{\approx}{\Delta}_0$$

$$= \frac{1}{I}\left[\sum_{i=0}^{I}\binom{I}{i}\left(\frac{1}{n}\right)^i\left(\frac{n-1}{n}\right)^{I-i}\right]\underset{\approx}{\Delta}_0$$

$$= \frac{1}{I}\cdot\underset{\approx}{\Delta}_0 > \underset{\sim}{0}.$$

Note that $\frac{1}{I}\cdot\underset{\approx}{\Delta}_0$ is of order $\boldsymbol{O(1)}$ since each $\underset{\approx}{\Delta}_l$ is $\boldsymbol{O(I)}$. Therefore, we have shown that $Var\left(\frac{1}{Q\sqrt{I}}\sum_{q}^{Q}\sum_{i}^{I}\boldsymbol{U_i}(\underset{\sim}{\boldsymbol{\beta}};\boldsymbol{q})\right)$ goes to a non-zero finite limit at least of order $\boldsymbol{O(1)}$.

This argument is easily extendible to situations where $n_i \neq n\ \forall i$.

This approach to showing the asymptotic normality ignores the dependency of the observations within a cluster. We have merely established asymptotic normality with an unspecified, but finite covariance matrix, $\underset{\sim}{\Sigma}$.

3.5. WCR Variance Estimator

In the previous section, the asymptotic normality of a standardized version of the WCR estimator was established. However, a consistent estimator for the variance needs to be determined. The following proposition defines and establishes a consistent WCR variance estimator.

Proposition 4: Let I and $\underset{\sim}{\overline{\beta}}$ be defined as in Proposition 3 of Section 3.4. Let us define

$$\underset{\sim}{\widehat{\Sigma}} = \widehat{Var}(\sqrt{I}(\underset{\sim}{\overline{\beta}} - \underset{\sim}{\beta})) \doteq I \cdot \left(\frac{\sum_{q=1}^{Q} \underset{\sim}{\widehat{\Sigma}}(R;q)}{Q} - (\tfrac{Q-1}{Q}) \cdot \underset{\sim}{S}^{2}_{\underset{\sim}{\beta}} \right),$$

where $\underset{\sim}{\widehat{\Sigma}}(R;q)$ is the estimated covariance matrix from the q^{th} analysis and $\underset{\sim}{S}^{2}_{\underset{\sim}{\beta}} = \frac{\sum_{q=1}^{Q} (\underset{\sim}{\widehat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}})(\underset{\sim}{\widehat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}})^T}{Q-1}$ is the estimated covariance matrix among the $\underset{\sim}{\widehat{\beta}}(R;q)$, $q = 1, \ldots, Q$. Then $\underset{\sim}{\widehat{\Sigma}}$ is a consistent estimator of $\underset{\sim}{\Sigma} = Var(\sqrt{I}(\underset{\sim}{\overline{\beta}} - \underset{\sim}{\beta}))$.

Proof:

If the iterated variance is applied to a single resample-based estimator, $\underset{\sim}{\widehat{\beta}}(R)$, then

$$Var(\sqrt{I} \cdot \underset{\sim}{\widehat{\beta}}(R)) = E\{var(\sqrt{I} \cdot \underset{\sim}{\widehat{\beta}}(R))|data\} + Var\{E(\sqrt{I} \cdot \underset{\sim}{\widehat{\beta}}(R))|data\}, \quad (3.35)$$

where the terms on the right hand side of (3.35) are the respective expectation and variance over the Q resample-based data sets. This variance representation can be simplified by noticing that the last term on the right-hand side of (3.35) is the variance of $\overline{\underset{\sim}{\beta}}$ since $E(\widehat{\underset{\sim}{\beta}}\,(R)|data) = \overline{\underset{\sim}{\beta}}$,

$$Var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R)) = E\{var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R))|data\} + Var(\sqrt{I}\cdot\overline{\underset{\sim}{\beta}}\,). \tag{3.36}$$

Rearranging (3.36), we get

$$Var(\sqrt{I}\cdot\overline{\underset{\sim}{\beta}}\,) = Var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R)) - E\{var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R))|data\}. \tag{3.37}$$

The first term on the right hand side of (3.37), is the true unconditional variance of a single resample-based estimator. Each $I\cdot\widehat{\underset{\sim}{\Sigma}}\,(\boldsymbol{R})$, where $\widehat{\underset{\sim}{\Sigma}}\,(\boldsymbol{R})$ is the estimated variance matrix for a random resampled data set, is a consistent estimator of $Var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R))$. Therefore, the average of Q consistent estimators, $\frac{I\cdot\sum_q\widehat{\underset{\sim}{\Sigma}}(\boldsymbol{R};\boldsymbol{q})}{Q}$, will be a consistent estimator of $Var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R))$.

The second part on the right hand side of (3.37) is

$$E\{var(\sqrt{I}\cdot\widehat{\underset{\sim}{\beta}}\,(R))|data\} = \frac{I\cdot E(\sum_{q=1}^{Q}(\widehat{\underset{\sim}{\beta}}(R;q)-\overline{\underset{\sim}{\beta}})(\widehat{\underset{\sim}{\beta}}(R;q)-\overline{\underset{\sim}{\beta}})^T)}{Q} = \left(\frac{Q-1}{Q}\right)E(I\cdot\underset{\sim}{\boldsymbol{S}}^{\,2}_{\underset{\sim}{\beta}}), \tag{3.38}$$

which is exactly true when $Q\to\infty$.

It remains to show that

$$I(\underset{\sim}{\boldsymbol{S}}^{\,2}_{\underset{\sim}{\beta}} - E(\underset{\sim}{\boldsymbol{S}}^{\,2}_{\underset{\sim}{\beta}})) \xrightarrow{p} \underset{\sim}{0}\,. \tag{3.39}$$

An approach for showing (3.39) is to verify that $Var(I \cdot \underset{\sim}{S}^2_{\underset{\sim}{\beta}}) \to 0$, which comes from Markov's Theorem (Gnedenko, 1962):

> ***Markov's Theorem:***
>
> *If the sequence of random variables* $\xi_1, \xi_2, \ldots, \xi_n, \ldots,$ *is such that* $\frac{1}{n^2} Var(\sum_{k=1}^{n} \xi_k) \to 0$ *for* $n \to \infty$, *then for any positive constant* ε, $\lim_{n\to\infty} Pr\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E(\xi_k)\right| < \varepsilon\} = 1.$

Following this line of argument, it suffices to show that $Var(I \cdot \underset{\sim}{S}^2_{\underset{\sim}{\beta}}) \to 0$

$$Var(I \cdot \underset{\sim}{S}^2_{\underset{\sim}{\beta}}) = Var\left(\frac{\sum_q (\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T)}{Q}\right)$$

$$= \frac{\sum_q Var[\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T]}{Q^2}$$
$$+ \frac{\sum\sum_{q \neq q'} Cov[(\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T), (\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta})^T)]}{Q^2}$$

$$\doteq \frac{Var[\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T]}{Q}$$
$$+ \frac{Q(Q-1) Cov[(\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T), (\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta}) \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta})^T)]}{Q^2}$$

$$\doteq O(\tfrac{1}{Q})$$
$$+ I^2 \cdot Cov[(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot (\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T, (\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta}) \cdot (\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta})^T]$$

$$= o_p(1)$$
$$+ I^2 \cdot Cov[(\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta}) \cdot (\hat{\underset{\sim}{\beta}}(R;q) - \underset{\sim}{\beta})^T, (\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta}) \cdot (\hat{\underset{\sim}{\beta}}(R;q') - \underset{\sim}{\beta})^T]$$

$$\doteq I^2 \cdot Cov[(\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,) \cdot (\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,)^T, (\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,) \cdot (\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,)^T]. \quad (3.40)$$

It therefore suffices to show that the expression in (3.40) goes to zero as $I \rightarrow \infty$. The proof proceeds by the following steps:

1. Proving the asymptotic normality of $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R) - \overline{\underset{\sim}{\beta}}\,)$ as $I \rightarrow \infty$;

2. Show $Cov(\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,), \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,)) \rightarrow 0$ for two random resamples;

3. Prove bivariate normality of $(\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,), \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,))$ as $I \rightarrow \infty$;

4. Show $I^2 \cdot Cov[(\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,) \cdot (\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,)^T, (\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,) \cdot (\widehat{\underset{\sim}{\beta}}\,(R;q') - \overline{\underset{\sim}{\beta}}\,)^T] \rightarrow 0.$

We begin by proving the asymptotic normality of $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R) - \overline{\underset{\sim}{\beta}}\,)$ which holds as $I \rightarrow \infty$. The argument directly parallels the proof of Proposition 3. Note the following

$$\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R) - \overline{\underset{\sim}{\beta}}\,) = \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R) - \underset{\sim}{\beta}\,) - \sqrt{I}(\overline{\underset{\sim}{\beta}} - \underset{\sim}{\beta}\,). \quad (3.41)$$

By Propositions 1 and 3, equation (3.41) can be approximated by

$$\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q) - \overline{\underset{\sim}{\beta}}\,) \doteq \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \boldsymbol{U_I}(\underset{\sim}{\boldsymbol{\beta}}\,;\boldsymbol{q}) + \boldsymbol{o(1)} \text{ in the } 1^{st} \text{ mean}$$
$$- \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \sum_{q'=1}^{Q} \boldsymbol{U_I}(\underset{\sim}{\boldsymbol{\beta}}\,;\boldsymbol{q'}) - \frac{1}{Q}\sum_{q'} \boldsymbol{o(1)} \text{ in the } 1^{st} \text{ mean}$$

$$= \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \underset{\sim}{U}_I(\underset{\sim}{\beta}\,; q) - \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \sum_{q'=1}^{Q} \underset{\sim}{U}_I(\underset{\sim}{\beta}\,; q') + o(1) \text{ in the } 1^{st} \text{ mean,}$$

where $\frac{1}{Q}\sum_{q'} o(1)$ in the 1^{st} mean $= o(1)$ in the 1^{st} mean by condition (ii). Continuing, we get

$$\doteq \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left\{ \frac{(Q-1)\underset{\sim}{U}_I(\underset{\sim}{\beta};q) - \sum_{q' \neq q}^{Q} \underset{\sim}{U}_I(\underset{\sim}{\beta};q')}{Q} \right\}$$

$$= \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left\{ \frac{\sum_{i=1}^{I}(Q-1)U_{Ii}(\underset{\sim}{\beta};q) - \sum_{q' \neq q}^{Q}\sum_{i=1}^{I} U_{Ii}(\underset{\sim}{\beta};q')}{Q} \right\}$$

$$= \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left\{ \frac{\sum_{i=1}^{I}\{(Q-1)U_{Ii}(\underset{\sim}{\beta};q) - \sum_{q' \neq q}^{Q} U_{Ii}(\underset{\sim}{\beta};q')\}}{Q} \right\}$$

$$= \frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left(\frac{\sum_{i=1}^{I} U_{Ii}^{(q)}(\underset{\sim}{\beta})}{Q} \right), \tag{3.42}$$

where $\underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta}) = \{(Q-1)\underset{\sim}{U}_{Ii}(\underset{\sim}{\beta}\,; q) - \sum_{q' \neq q}^{Q} \underset{\sim}{U}_{Ii}(\underset{\sim}{\beta}\,; q')\}$. It should be clear that each $\frac{U_{Ii}^{(q)}(\underset{\sim}{\beta})}{Q}$ is $O_p(1)$ and $\sum_{i=1}^{I} \underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta})$ is the sum of independent and identically distributed random variables. Therefore, the Central Limit Theorem may be directly applied to (3.42). Using Slutsky's Theorem

$$\sqrt{I}(\underset{\sim}{\hat{\beta}}\,(R) - \underset{\sim}{\overline{\beta}}) \sim \underset{\sim}{N}_p(\underset{\sim}{0}, \underset{\sim}{\Gamma}), \text{ as } I \to \infty \tag{3.43}$$

where $\underset{\sim}{\Gamma}$ is a finite, positive-definite matrix, as shown in the following proof.

Proof:

$$\underset{\sim}{\Gamma} = Var(\sqrt{I}(\underset{\sim}{\hat{\beta}}(R) - \underset{\sim}{\bar{\beta}})) = I \cdot Var(\underset{\sim}{\hat{\beta}}(R) - \underset{\sim}{\bar{\beta}})$$

$$= I \cdot Var(\tfrac{Q-1}{Q}\underset{\sim}{\hat{\beta}}(R) - \tfrac{1}{Q}\sum_{q \neq q'}^{Q}\underset{\sim}{\hat{\beta}}(R;q))$$

$$= I[Var(\tfrac{Q-1}{Q}\underset{\sim}{\hat{\beta}}(R)) + Var(\tfrac{1}{Q}\sum_{q \neq q'}^{Q}\underset{\sim}{\hat{\beta}}(R;q)) - 2Cov(\tfrac{Q-1}{Q}\underset{\sim}{\hat{\beta}}(R), \tfrac{1}{Q}\sum_{q \neq q'}^{Q}\underset{\sim}{\hat{\beta}}(R;q))]$$

$$= I[(\tfrac{Q-1}{Q})^2 Var(\underset{\sim}{\hat{\beta}}(R)) + \tfrac{(Q-1)}{Q^2}Var(\underset{\sim}{\hat{\beta}}(R)) + \tfrac{(Q-1)(Q-2)}{Q^2}Cov(\underset{\sim}{\hat{\beta}}(R;q), \underset{\sim}{\hat{\beta}}(R;q')) - \tfrac{2(Q-1)^2}{Q^2}Cov(\underset{\sim}{\hat{\beta}}(R;q), \underset{\sim}{\hat{\beta}}(R;q'))]$$

$$\doteq I[Var(\underset{\sim}{\hat{\beta}}(R)) + O(\tfrac{1}{Q}) - Cov(\underset{\sim}{\hat{\beta}}(R;q), \underset{\sim}{\hat{\beta}}(R;q'))]$$

$$\doteq I[Var(\underset{\sim}{\hat{\beta}}(R)) - Cov(\underset{\sim}{\hat{\beta}}(R;q), \underset{\sim}{\hat{\beta}}(R;q'))] \geq \underset{\sim}{0} \text{ as } Q \to \infty. \qquad (3.44)$$

by (3.10) and the Cauchy-Schwarz Inequality. Thus, we have established the asymptotic normality of $\sqrt{I}(\underset{\sim}{\hat{\beta}}(R) - \underset{\sim}{\bar{\beta}})$ and shown that it has a finite, positive variance.

Next we will prove that $Cov(\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\bar{\beta}}), \sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q') - \underset{\sim}{\bar{\beta}})) \to \underset{\sim}{0}$. Note that when an average equals zero, such as $\frac{1}{Q}\sum_{q=1}^{Q}\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\bar{\beta}}) = \underset{\sim}{0}$, then the variance of this average also equals zero, $Var(\frac{1}{Q}\sum_{q=1}^{Q}\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\bar{\beta}})) = \underset{\sim}{0}$. It follows from the exchangeability of the estimates, $\underset{\sim}{\hat{\beta}}(R)$, that the covariance also is zero by the following argument

$$Var(\tfrac{1}{Q}\sum_{q=1}^{Q}\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,)) = \tfrac{I}{Q^2}\sum_{q=1}^{Q}Var(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,)$$
$$+\tfrac{IQ(Q-1)}{Q^2}Cov[(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,),(\widehat{\underset{\sim}{\beta}}\,(R;q')-\overline{\underset{\sim}{\beta}}\,)] = \underset{\sim}{0}\,.$$

As $Q \to \infty$, we have

$$I \cdot Cov[(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,),(\widehat{\underset{\sim}{\beta}}\,(R;q')-\overline{\underset{\sim}{\beta}}\,)] = -\tfrac{I}{Q-1}Var(\widehat{\underset{\sim}{\beta}}\,(R)-\overline{\underset{\sim}{\beta}}\,) \Rightarrow \underset{\sim}{0}\,. \quad (3.45)$$

Now we must show that $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,)$ and $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q')-\overline{\underset{\sim}{\beta}}\,)$ are asymptotically bivariate normal. We use Theorem 2.5 of Seber (1977):

> ***Theorem 2.5:***
>
> *$\boldsymbol{Y}$ has a multivariate normal distribution if and only if $\boldsymbol{a}^T\boldsymbol{Y}$ is univariate normal for all real vectors $\boldsymbol{a}$ ($\boldsymbol{a} \neq \boldsymbol{0}$).*

Proof:

We know that $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R)-\overline{\underset{\sim}{\beta}}\,) \sim \boldsymbol{N_p}(\underset{\sim}{0}\,,\underset{\sim}{\Gamma}\,)$, as $I \to \infty$, therefore $\underset{\sim}{C}_q \cdot \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,)$ is $\boldsymbol{N_d}(\underset{\sim}{0}\,,\underset{\sim}{C}_q\underset{\sim}{\Gamma}\,\underset{\sim}{C}_q^T)$ where $\underset{\sim}{C}_q$ is a $(d \times p)$ matrix (Seber, 1977). Define $\underset{\sim}{C}\underset{\sim}{V} = \underset{\sim}{C}_q \cdot \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,) + \underset{\sim}{C}_{q'} \cdot \sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q')-\overline{\underset{\sim}{\beta}}\,)$, where $\underset{\sim}{C} = (\underset{\sim}{C}_q, \underset{\sim}{C}_{q'})$ and specifically $\underset{\sim}{C}_q$ and $\underset{\sim}{C}_{q'}$ are $(1 \times p)$ matrices so $\underset{\sim}{C}\underset{\sim}{V}$ is a (1×1) matrix of an arbitrary linear combination of the elements of $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q)-\overline{\underset{\sim}{\beta}}\,)$ and $\sqrt{I}(\widehat{\underset{\sim}{\beta}}\,(R;q')-\overline{\underset{\sim}{\beta}}\,)$. We write

$$\underset{\sim}{C}\underset{\sim}{V} = \underset{\sim}{C}_q \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}}) + \underset{\sim}{C}_{q'} \cdot \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \overline{\underset{\sim}{\beta}})$$

$$\doteq \underset{\sim}{C}_q \left(\frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left(\frac{\sum_{i=1}^{I} U_{Ii}^{(q)}(\underset{\sim}{\beta})}{Q} \right) \right) + \underset{\sim}{C}_{q'} \left(\frac{J^{-1}(\underset{\sim}{\beta})}{\sqrt{I}} \left(\frac{\sum_{i=1}^{I} U_{Ii}^{(q')}(\underset{\sim}{\beta})}{Q} \right) \right)$$

$$= \underset{\sim}{C}_q \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \sum_{i=1}^{I} \underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta}) + \underset{\sim}{C}_{q'} \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \sum_{i=1}^{I} \underset{\sim}{U}_{Ii}^{(q')}(\underset{\sim}{\beta})$$

$$= \sum_{i=1}^{I} \{ \underset{\sim}{C}_q \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta}) + \underset{\sim}{C}_{q'} \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \underset{\sim}{U}_{Ii}^{(q')}(\underset{\sim}{\beta}) \}$$

$$= \sum_{i=1}^{I} V_i(\underset{\sim}{\beta}; q, q'), \tag{3.46}$$

where $\underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta}) = \{(Q-1)U_{Ii}(\underset{\sim}{\beta}; q) - \sum_{r \neq q}^{Q} U_{Ii}(\underset{\sim}{\beta}; r)\}$ and $V_i(\underset{\sim}{\beta}; q, q') = \underset{\sim}{C}_q \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \underset{\sim}{U}_{Ii}^{(q)}(\underset{\sim}{\beta}) + \underset{\sim}{C}_{q'} \frac{J^{-1}(\underset{\sim}{\beta})}{Q\sqrt{I}} \underset{\sim}{U}_{Ii}^{(q')}(\underset{\sim}{\beta})$. Therefore, since (3.46) can be written as a sum of independent random variables we can apply the Central Limit Theorem. As in (3.42), we can claim asymptotic univariate normality of $\underset{\sim}{C}\underset{\sim}{V} \sim N(\underset{\sim}{0}, \underset{\sim}{C}\underset{\sim}{\Gamma}^{*}\underset{\sim}{C})$ where $\underset{\sim}{\Gamma}^{*} = \begin{pmatrix} \underset{\sim}{\Gamma} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Gamma} \end{pmatrix}$ and $Var(\underset{\sim}{C}\underset{\sim}{V}) = \underset{\sim}{C}\underset{\sim}{\Gamma}^{*}\underset{\sim}{C}$ because $Cov(\sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}}), \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \overline{\underset{\sim}{\beta}})) \rightarrow \underset{\sim}{0}$ as shown in (3.45). Again applying Slutsky's Theorem as in (3.43), we can conclude

$$\begin{pmatrix} \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}}) \\ \sqrt{I}(\hat{\underset{\sim}{\beta}}(R;q') - \overline{\underset{\sim}{\beta}}) \end{pmatrix} \sim \underset{\sim}{N}_{2p}(\underset{\sim}{0}, \underset{\sim}{\Gamma}^{*}), \text{ as } I \rightarrow \infty \text{ and } Q \rightarrow \infty. \tag{3.47}$$

Now that we have established the bivariate normality as in (3.47), we can use some well-known properties of the bivariate normal distribution to demonstrate

$$I^2 \cdot Cov[(\hat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}}) \cdot (\hat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}})^T, (\hat{\underset{\sim}{\beta}}(R;q') - \overline{\underset{\sim}{\beta}}) \cdot (\hat{\underset{\sim}{\beta}}(R;q') - \overline{\underset{\sim}{\beta}})^T] \rightarrow 0. \tag{3.48}$$

Specifically, we define $\underset{\sim}{X} = \underset{\sim}{\Gamma}^{-1/2}(\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}}))$,

$\underset{\sim}{Y} = \underset{\sim}{\Gamma}^{-1/2}(\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q') - \underset{\sim}{\overline{\beta}}))$, and $\underset{\sim}{Z} = \underset{\sim}{Y} - \underset{\sim}{\Phi}\,\underset{\sim}{X}$, where

$\underset{\sim}{\Phi} = Corr(\underset{\sim}{X}, \underset{\sim}{Y}) = \underset{\sim}{\Gamma}^{-1} \cdot \underset{\sim}{\Psi}$ and

$\underset{\sim}{\Psi} = Corr(\sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}}), \sqrt{I}(\underset{\sim}{\hat{\beta}}(R;q') - \underset{\sim}{\overline{\beta}}))$. This implies

$E(\underset{\sim}{X}) = E(\underset{\sim}{Y}) = \underset{\sim}{0}$, $V(\underset{\sim}{X}) = V(\underset{\sim}{Y}) = \underset{\sim}{I}_p$, where $\underset{\sim}{I}_p$ is a $(p \times p)$ identity matrix,

$E(\underset{\sim}{X}\,\underset{\sim}{X}^T) = E(\underset{\sim}{Y}\,\underset{\sim}{Y}^T) = \underset{\sim}{I}_p$,

$E(\underset{\sim}{X}\,\underset{\sim}{X}^T\,\underset{\sim}{X}) = E(\underset{\sim}{Y}\,\underset{\sim}{Y}^T\,\underset{\sim}{Y}) = \underset{\sim}{0}$, $E\{(\underset{\sim}{X}\,\underset{\sim}{X}^T)(\underset{\sim}{X}\,\underset{\sim}{X}^T)^T\} =$

$E\{(\underset{\sim}{Y}\,\underset{\sim}{Y}^T)(\underset{\sim}{Y}\,\underset{\sim}{Y}^T)^T\} = 3 \cdot \underset{\sim}{I}_p$ and $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ are independent. We begin by writing

$$I^2 \cdot Cov[(\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}}) \cdot (\underset{\sim}{\hat{\beta}}(R;q) - \underset{\sim}{\overline{\beta}})^T, (\underset{\sim}{\hat{\beta}}(R;q') - \underset{\sim}{\overline{\beta}}) \cdot (\underset{\sim}{\hat{\beta}}(R;q') - \underset{\sim}{\overline{\beta}})^T]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot Cov(\underset{\sim}{X}\,\underset{\sim}{X}^T, \underset{\sim}{Y}\,\underset{\sim}{Y}^T)$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [E(\underset{\sim}{X}\,\underset{\sim}{X}^T\,\underset{\sim}{Y}\,\underset{\sim}{Y}^T) - E(\underset{\sim}{X}\,\underset{\sim}{X}^T)E(\underset{\sim}{Y}\,\underset{\sim}{Y}^T)]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [E(\underset{\sim}{X}\,\underset{\sim}{X}^T\,\underset{\sim}{Y}\,\underset{\sim}{Y}^T) - \underset{\sim}{I}_p]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [E(\underset{\sim}{X}\,\underset{\sim}{X}^T(\underset{\sim}{Z} + \underset{\sim}{\Phi}\,\underset{\sim}{X})(\underset{\sim}{Z} + \underset{\sim}{\Phi}\,\underset{\sim}{X})^T - \underset{\sim}{I}_p]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [E(\underset{\sim}{X}\,\underset{\sim}{X}^T\,\underset{\sim}{Z}\,\underset{\sim}{Z}^T + 2\underset{\sim}{X}\,\underset{\sim}{X}^T\underset{\sim}{\Phi}\,\underset{\sim}{X}\,\underset{\sim}{Z}^T + \underset{\sim}{\Phi}\,\underset{\sim}{\Phi}^T\underset{\sim}{X}\,\underset{\sim}{X}^T(\underset{\sim}{X}\,\underset{\sim}{X}^T)^T) - \underset{\sim}{I}_p]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [\underset{\sim}{I}_p - \underset{\sim}{\Phi}\,\underset{\sim}{\Phi}^T + 0 + (3 \cdot \underset{\sim}{I}_p)\underset{\sim}{\Phi}\,\underset{\sim}{\Phi}^T - \underset{\sim}{I}_p]$$

$$= (\underset{\sim}{\Gamma}\,\underset{\sim}{\Gamma}^T) \cdot [\underset{\sim}{I}_p + 2\underset{\sim}{\Phi}\,\underset{\sim}{\Phi}^T - \underset{\sim}{I}_p] = 2\underset{\sim}{\Psi}\,\underset{\sim}{\Psi}^T = O(\tfrac{1}{Q^2}) \to \underset{\sim}{0}.$$

Therefore, $Var(I \cdot \underset{\sim}{S}^2_{\underset{\sim}{\beta}})$ is $O(\frac{1}{Q})$ and goes to zero very quickly since Q is very large. Consequently, a consistent estimate of $\underset{\sim}{\Sigma}$ is

$$\widehat{\underset{\sim}{\Sigma}} = I \cdot \widehat{Var}(\underset{\sim}{\beta}) \doteq I \cdot \left(\frac{\sum_{q=1}^{Q} \widehat{\underset{\sim}{\Sigma}}(R;q)}{Q} - (\frac{Q-1}{Q}) \cdot \underset{\sim}{S}^2_{\underset{\sim}{\beta}} \right), \tag{3.49}$$

where $\widehat{\underset{\sim}{\Sigma}}\ (R;q)$ is the estimated covariance matrix from the q^{th} analysis and

$$\underset{\sim}{S}^2_{\underset{\sim}{\beta}} = \frac{\sum_{q=1}^{Q} (\widehat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}})(\widehat{\underset{\sim}{\beta}}(R;q) - \overline{\underset{\sim}{\beta}})^T}{Q-1} \tag{3.50}$$

is the covariance matrix among $\widehat{\underset{\sim}{\beta}}\ (R;q)$'s.

3.6. Examples

The first example uses data from Chen, Kodell, Howe, and Gaylor (1991). The data are from a study on the developmental effects of fetuses resulting from exposure of pregnant dams to hydroxyurea. The experiment consisted of three doses of hydroxyurea: low, medium, and high. The fetuses of the 90 dams were observed for malformation and death, however, the following analyses will only model the probability of death. The cluster size ranged from 5 to 16 fetuses within a litter. The data are as follows:

Table 3.2. Data from Chen, et al. (1991)

Low	1	0	0	5	1	0	2	2	1	0	2	1	0	4	0	1	2	1	0	1	1	0	1	
	14	11	11	8	12	8	11	10	12	14	13	7	9	9	11	14	14	10	10	9	10	9	10	
Medium	2	0	12	9	12	5	3	3	10	4	3	9	9	8	4	2	1	8	2	4				
	14	11	16	11	13	13	13	11	12	13	9	10	11	13	8	10	12	11	12	12				
High	7	8	6	10	7	3	4	9	7	4	11	5	10	10	6	8	6	8	11	2	11	2		
	9	11	11	11	11	11	9	10	8	5	12	12	12	11	11	13	12	9	12	10	12	9		
	1	7	10	9	10	12	4	3	11	7	10	15	1	4	5	7	4	2	4	3	11	7		
	11	10	12	11	12	14	12	9	12	12	12	16	9	9	9	11	12	10	11	11	12	10		

*Within each treatment group the top number is the number of dead fetuses and the number below it is the total number of fetuses in the litter.

The Chen, et al. data set will help illustrate the distinct difference in definition and interpretation between the traditional marginal parameter and the WCR marginal parameter. The traditional marginal parameter is the probability that a *randomly selected fetus* is abnormal. Using the low treatment group from the above data set, the traditional marginal risk is computed as

$$\widehat{Pr}_{(\mathrm{TM})}(\text{abnormal}) = \frac{\text{total number of abnormal fetuses}}{\text{total number of fetuses}}$$

$$= \frac{(1+0+0+5+1+0+2+2+1+0+2+1+0+4+0+1+2+1+0+1+1+0+1)}{(14+11+11+8+12+8+11+10+12+14+13+7+9+9+11+14+14+10+10+9+10+9+10)}$$

$$= 0.1057.$$

The marginal risk from WCR is the probability that a *randomly selected fetus from a randomly selected litter* is abnormal. Again using the low treatment group, the marginal risk from WCR is computed as

$$\widehat{Pr}_{(\text{WCR})}(\text{abnormal}) = \frac{\text{sum of litter-specific abnormality } \hat{p}\text{'s}}{\text{number of litters}}$$

$$= \frac{(\frac{1}{14}+\frac{0}{11}+\frac{0}{11}+\frac{5}{8}+\frac{1}{12}+\frac{0}{8}+\frac{2}{11}+\frac{2}{10}+\frac{1}{12}+\frac{0}{14}+\frac{2}{13}+\frac{1}{7}+\frac{0}{9}+\frac{4}{9}+\frac{0}{11}+\frac{1}{14}+\frac{2}{14}+\frac{1}{10}+\frac{0}{10}+\frac{1}{9}+\frac{1}{10}+\frac{0}{9}+\frac{1}{10})}{23}$$

$$= 0.1135.$$

Note that the difference is in the weighting of the clusters. WCR *equally weights* each litter, whereas, the traditional marginal approach *increases the weights with the size* of the litters.

The WCR, GEE, and OLR methods were used to analyze the Chen, et al. (1991) data set. All three methods fit the same logistic model:

$$ln\left(\frac{\pi_k}{1-\pi_k}\right) = \beta_0 + \beta_1 x \tag{3.51}$$

where π_k is the probability of death for a fetus in the k^{th} treatment group and x is the dichotomous treatment variable: $x = 0$ if low or medium; $x = 1$ if high hydroxyurea group. For the WCR analysis this model was fit in each of the $Q = 10,000$ resampled data sets using Proc GENMOD in SAS. Each of the three analysis methods treats the within-cluster correlation differently. The WCR analysis takes into account the dependency through the variance formula in (3.53), which imposes no particular form on the correlation structure. The GEE analysis relies upon a user-specified working correlation matrix, which for this analysis was chosen to be exchangeable. The OLR analysis assumes that all of the observations within a litter are independent so it proceeds as if each observation contributes 100% new information to the analysis, which is not true. Therefore, the variance estimates with positive within-cluster correlation from OLR tend to be underestimated, which results in inflated Type I error rates.

For finite samples, there is a small probability of getting infinite parameter estimates due to the nature of the WCR resampling procedure. When this situation arises, we only use the finite part of the resampling distribution for parameter estimation and variance calculations. In Chapter IV, simulations will show that the WCR analysis method works quite well in small samples using this technique.

The cluster-based and observation-based estimated probabilities of death of a fetus by treatment group were calculated to support the slight difference in parameter estimates produced by the three analysis methods. These estimates are given in Table 3.3.

Table 3.3. Estimated Probability of Death by Treatment Group

	Cluster-based	Observation-based
$x = 0$	0.2780	0.2827
$x = 1$	0.6353	0.6430

The analysis results are as follows:

Table 3.4. CHEN, ET AL. (1991) ANALYSIS RESULTS

	WCR	GEE(r)	GEE(mb)	OLR
Q	10,000			
I	90	90	90	90
n_i	5 – 16	5 – 16	5 – 16	5 – 16
$\widehat{\beta}_0$	– 0.9732	– 0.9309	– 0.9309	– 0.9309
$\widehat{\beta}_1$	1.5403	1.5193	1.5193	1.5193
$var(\widehat{\beta}_0)$	0.0499	0.0502	0.0404	0.0102
$var(\widehat{\beta}_1)$	0.0761	0.0773	0.0745	0.0189
Z_1	5.58	5.46	5.57	11.07

*GEE(r) is the robust variance; GEE(mb) is the model-based variance.

It is interesting to note that the parameter estimates are consistent with the probabilities found in Table 3.3, which told us to expect the WCR intercept parameter estimate to be slightly smaller than the GEE and OLR intercept parameter estimates. Although all of the approaches produce nearly the same parameter estimates, it is clear that OLR is not accounting for the within-cluster correlation since the variance estimates are much too small. The WCR analysis method is accounting for the within-cluster correlation which is evident by the variance estimates being quite close in value to the GEE robust-variance counterparts.

The second data analysis uses a continuous covariate. The data are from a mouse teratology experiment first published by Williams (1988b). The litter sizes range from 3 to 19 and the number of deaths per litter are counted. The continuous covariate is the dosage in g/kg of body weight and has values of 0, 0.75, 1.5, and 3.0. The Williams (1988b) data are as follows:

Table 3.5. Data from Williams (1988b)

Dosage																										
0.00	1	2	0	1	1	5	0	0	1	0	0	1	3	0	2	0	1	2	2	2	5	0	2	2	3	
	9	9	11	11	11	11	12	12	12	13	13	13	13	14	14	15	15	15	15	15	15	16	16	16	17	
0.75	0	2	0	0	1	2	0	1	1	0	1	1	2	5	0	1	2	4	0	0	0	1	5	5		
	7	9	10	11	11	11	12	12	12	13	13	13	13	13	14	14	14	14	15	15	15	15	15	19		
1.50	3	4	4	2	2	0	0	1	1	2	3	0	0	1	2	1	2	3	0	2	2	2				
	3	5	7	9	9	10	12	12	12	12	12	13	13	13	13	14	14	14	15	15	15	16				
3.00	2	2	5	1	6	2	0	0	1	1	2	2	2	5	1	1	1	2	4	3	3	2	9			
	9	9	9	10	10	11	12	12	12	12	12	12	12	12	13	13	13	14	14	15	15	16	16			

*Within each treatment group the top number is the number of dead fetuses and the number below it is the total number of fetuses in the litter.

A logistic model, similar to (3.51), was fit to each of $Q = 5000$ resampled data sets in the WCR analysis, as well as in GEE and OLR with dosage as a continuous covariate. Again an exchangeable correlation structure was specified for the GEE analysis. The analysis results are in Table 3.6.

Table 3.6. WILLIAMS (1988b) ANALYSIS RESULTS

	WCR	GEE(r)	GEE(mb)	OLR
Q	5000			
I	94	94	94	94
n_i	3 – 19	3 – 19	3 – 19	3 – 19
$\widehat{\beta}_0$	– 2.1548	– 2.1877	– 2.1877	– 2.1877
$\widehat{\beta}_1$	0.2808	0.2645	0.2645	0.2645
$var(\widehat{\beta}_0)$	0.0326	0.0324	0.0396	0.0189
$var(\widehat{\beta}_1)$	0.0098	0.0102	0.0112	0.0053
Z_1	2.83	2.62	2.51	3.62

*GEE(r) is the robust variance; GEE(mb) is the model-based variance.

Again, the parameter estimates from the three analysis methods are quite similar. It is clear that the WCR analysis method is accounting for the within-cluster correlation as the variance estimates are virtually identical to the GEE robust variance estimates.

3.7. Summary

The WCR analysis method is a valid way to analyze clustered data. It is a marginal analysis method with a cluster-based parameter. The WCR parameter is different than the traditional marginal parameter because of the WCR sampling scheme. However, the value and interpretation of the WCR marginal parameter and the traditional marginal parameter may be different due to the weighting of the clusters. In a litter-effects study, dams who naturally produce large litters may be at higher risk for producing pups with birth defects. Therefore, analysis methods that increase the weight

with the size of the cluster overweight this portion of the study population. WCR gives equal weight to all clusters so the cluster size does not affect the analysis results.

The WCR estimator, the average of the Q resample-based estimates, was shown to asymptotically normal by rewriting the score statistic as a sum of independent random variables. The consistent WCR variance estimator is quite intuitive. The variance is the difference of two quantities; the first of which is the average of the estimated variance matrix from each of the Q resample-based analyses. The second quantity is the variability among the resampled-based estimates. When there is zero within-cluster correlation, there should be great variability among the resample-based estimates. Therefore, the second quantity will be large and the WCR variance will be small. This is consistent with what is found with analyzing data by OLR. When the within-cluster correlation is nearly one, there should be little to no variability among the resample-based estimates. Thus, the second quantity will be nearly zero and the WCR variance will consist of only the first quantity. The first quantity would be the average of the same variance since each resampled data set is virtually identical. In Chapter IV, simulations were done using the logistic model that show that the WCR variance estimator works for a wide range of correlations, and has good coverage and power as compared with GEE.

CHAPTER IV

OPERATING CHARACTERISTICS OF WCR AND GEE

4.1 Introduction

Simulations are a computational means of validating the finite sample behavior of a method by performing a method a large number of times. Through simulations, the operating characteristics of a method can be shown to be acceptable or not, for moderate samples sizes. To demonstrate that WCR is a valid method for analyzing clustered binary data, various scenarios that reflect realistic situations were analyzed in an extensive simulation study of clustered binary data.

The simulations are patterned after studies of reproductive toxicology, where pregnant dams or the fetuses of pregnant dams are exposed or unexposed to some toxin. When a pregnant dam is exposed, this is a cluster-specific exposure. However, if the individual fetuses of a pregnant dam have different exposure status within a litter, then these are observation-specific exposures. The sex of the pup would be an example of an observation-specific covariate. Often the outcomes from litter-effect studies are the observed birth-defects, malformations, or death of the pups within the litters.

WCR is an intuitive method for analyzing litter-effects type data because of the sampling scheme. Some dams naturally produce larger litters than others and this variability in litter size can affect the parameter estimated in an analysis. The WCR method does not allow dams who naturally produce larger litters to contribute more to an analysis. All dams are given equal weight in a WCR analysis. This is in direct contrast to GEE, where big litters are weighted more. The weighting is important if the cluster size

is related to the baseline risk of the outcome, which we term "informative cluster sizes". More important than the weighting issue is that informative cluster sizes violate an implicit assumption of traditional marginal models.

4.1.1 Data Generation

Each of the following scenarios is a set of $S = 1000$ simulated data sets with varying amounts of within-cluster correlation as well as numbers of resamples and clusters. Each data set consists of I clusters. The clusters are of size n_i, where $i = 1, \ldots, I$. The data were generated and the WCR analyses were performed in C.

The baseline risk for an abnormal pup was set in the various scenarios to a commonly observed value in actual litter-effects studies. For zero within-cluster correlation or independent data, observations within a cluster were generated from uniform variates. Uniform variates were obtained from the Marsaglia random number generator implemented in C (see Appendix 2). Uniforms were compared to probability cut-offs to generate Bernoulli outcomes.

For non-zero within-cluster correlation, the cluster-specific baseline risks were generated from beta variates. One way to generate a $beta(a, b)$ variate with integer valued a and b is to find the minimum of $(a + b - 1)$ uniform variates (Rubinstein, 1981). For non-integer a and b, the following algorithm can be used to generate a $beta(a, b)$ variate (Rubinstein, 1981):

1. Generate U_1 and U_2 as $uniform(0, 1)$ variates
2. Let $Z_1 = U_1^{1/a}$ and $Z_2 = U_2^{1/b}$
3. If $Z_1 + Z_2 \geq 1$ then return to Step 1, otherwise go to Step 4
4. $W = \frac{Y_1}{Y_1 + Y_2} \sim beta(a, b)$

In this simulation study, the various scenarios had either equal or unequal, and if unequal either informative or non-informative cluster sizes within a set of simulated data sets. Informative cluster sizes will be discussed in greater detail in Chapter V. Unequal cluster sizes were obtained from truncated binomial distributed variates, discarding clusters of size zero. For each cluster, the value of the response for each of the n_i observations was found by comparing n_i unique uniform variates to an *a priori* defined baseline risk, μ, for independent data, or a cluster-specific $beta_{(i)}(a, b)$ variate for correlated data. If the uniform variate was less than or equal to the value compared, then the outcome for the unit was given the value 1, which represents an abnormality. However, if the uniform was greater than the value compared, then the outcome for the unit was assigned the value 0, which represents a normal pup. The independent data are binomial data and the correlated data are referred to as beta-binomial data. The name "beta-binomial" is appropriate because the number of abnormal observations within each cluster is distributed $binomial(u_i, n_i)$, and the cluster-specific probability of the outcome, u_i, is distributed as a $beta(a, b)$. Beta-binomial data imply an exchangeable correlation structure or equal correlation between all observations within a cluster. The within-cluster correlation is $\rho = \frac{1}{a+b+1}$.

The exposure status of the cluster (observation) is assigned by comparing another uniform variate with 0.5, so approximately half of the clusters (observations) are exposed.

The marginal risk (given WCR sampling) of an unexposed observation is the *a priori* baseline risk value, μ, for independent data or the mean of the beta distribution for correlated data. If Y_{ij}, the value of the j^{th} observation in the i^{th} cluster, is distributed $Bernoulli(u_i)$, where u_i is distributed $beta(a, b)$, then the marginal risk of the (i, j) unexposed unit is

$$E(Y_{ij}|\text{unexposed}) = E(E(Y_{ij}|u_i, \text{unexposed})) = E(u_i|\text{unexposed}) = \frac{a}{a+b}. \quad (4.1)$$

The exposed observations have an increased risk for the outcome of interest due to the addition of a fixed-effect on the logit scale. Setting the outcomes for exposed units require taking the logit of the cluster-specific $beta_{(i)}(a,b)$ variate, adding the exposure fixed-effect, taking the anti-logit of the sum, and comparing the resulting probability to distinct uniform variates for each exposed unit to determine the outcomes. To compute the marginal risk of an exposed observation we integrate out the cluster-specific baseline. Let Y_{ij} be the outcome for the (i,j) observation, the cluster-specific model can be written as

$$ln(\frac{E(Y_{ij}|u_i,\mathrm{E}_{ij})}{1-E(Y_{ij}|u_i,\mathrm{E}_{ij})}) = ln(\frac{u_i}{1-u_i}) + \alpha \mathrm{E}_{ij}, \tag{4.2}$$

where u_i is defined as a cluster-specific $beta_{(i)}(a,b)$ variate, E_{ij} is an indicator for exposure, and α is the within-cluster effect of exposure. Although the value of α is pre-determined in these simulations, this is not the true value of the marginal parameter associated with the exposure. The α is the parameter that would be estimated in a random-effects model. Typically with positive within-cluster correlation, the marginal parameter is closer to the null than the subject-specific parameter (Neuhaus and Jewell, 1990 ; Rosner and Milton, 1988). For correlated data, the WCR marginal risk for an exposed observation is obtained by performing the following integration

$$E(Y_{ij}|\text{exposed}) = \int_0^1 \frac{exp(\pi_i+\alpha)}{1+exp(\pi_i+\alpha)}(\frac{1}{B(a,b)})u_i^{a-1}(1-u_i)^{b-1}du_i. \tag{4.3}$$

where $\pi_i = ln(\frac{u_i}{1-u_i})$. For independent data, the marginal risk for exposed observations is set *a priori*.

The values of the marginal model parameters are obtained from the marginal risk of an unexposed and exposed observation. The marginal model equation is similar to the

subject-specific model equation (4.2), however, the parameters are population-averaged not cluster-specific in nature

$$ln(\frac{E_i[E_{j|i}(Y_{ij}|\mathrm{E}_{ij})]}{1-E_i[E_{j|i}(Y_{ij}|\mathrm{E}_{ij})]}) = ln(\frac{\mu}{1-\mu}) + \alpha^*\mathrm{E}_{ij}$$
$$= b_0 + b_1\mathrm{E}_{ij}. \qquad (4.4)$$

4.1.2 Program Provisions

Since each resampled data set is a set of independent observations with a dichotomous exposure variable, each resampled data set may be summarized by a 2x2 table. With small I, there is a positive probability of a zero cell in a 2x2 resampled data set table. Resamples with zero cells are deleted since infinite parameter and variance estimates result. Only the finite portion of the resampling distribution can be used for parameter estimation and variance calculations. A typical 2x2 table can be represented as follows

	$\overline{E}$	E	
$\overline{D}$	n_{00}	n_{01}	n_{0+}
D	n_{10}	n_{11}	n_{1+}
	n_{+0}	n_{+1}	I

where $\overline{E}$ represents unexposed observations, E represents exposed observations, $\overline{D}$ represents unaffected outcomes, and D represents affected outcomes. The n_{00} cell is the number of observations that are unaffected and unexposed, the n_{01} cell is the number of observations that are unaffected and exposed, the n_{10} cell is the number of observations that are affected and unexposed, and the n_{11} cell is the number of observations that are affected and exposed. The corresponding marginal totals are denoted with a '+' in place

of the subscript, and I is the total number of observations in the table as well as the number of clusters.

Large sample formulas were used to calculate the parameter estimates and variances from the 2x2 tables. A good rule to follow is that all of the cells should have at least 5 observations (Stokes, Davis, and Koch, 1995). The formulas used in the simulations to calculate the intercept, b_0, the variance of the intercept, $var(b_0)$, the exposure parameter, b_1, and the variance of the exposure parameter, $var(b_1)$ are as follows

$$b_0 = ln\left(\frac{n_{10}}{n_{00}}\right), \tag{4.5}$$

$$var(b_0) = \frac{1}{n_{10}} + \frac{1}{n_{00}}, \tag{4.6}$$

$$b_1 = ln\left(\frac{n_{00} \cdot n_{11}}{n_{01} \cdot n_{10}}\right), \tag{4.7}$$

$$var(b_1) = \frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} + \frac{1}{n_{11}}. \tag{4.8}$$

Formulas (4.5)-(4.8) show that resampled data sets with zero cells in a 2x2 table must be deleted since they produce infinities.

There is also a provision in the simulation programs to not allow negative variance estimates. The consistent WCR variance formula is

$$\widehat{\underset{\sim}{\Sigma}} = I \cdot \widehat{Var(\underset{\sim}{\beta})} \doteq I \cdot \left(\frac{\sum_{q=1}^{Q} \widehat{\underset{\sim}{\Sigma}}(R;q)}{Q} - \left(\frac{Q-1}{Q}\right) \cdot \underset{\sim}{S}^{2}_{\underset{\sim}{\beta}} \right) \tag{4.9}$$

where $\widehat{\underset{\sim}{\Sigma}}\ (R;q)$ is the estimated covariance matrix at the q^{th} analysis and

$$\underset{\sim}{S}^2_{\underset{\sim}{\beta}} = \frac{\sum_{q=1}^{Q} \underset{\sim}{\hat{\beta}}(R;q)\,\underset{\sim}{\hat{\beta}}(R;q)^T - Q\left(\frac{\sum_{q=1}^{Q} \underset{\sim}{\hat{\beta}}(R;q)}{Q}\right)\left(\frac{\sum_{q=1}^{Q} \underset{\sim}{\hat{\beta}}(R;q)}{Q}\right)^T}{Q-1}$$

is the variation among the $\underset{\sim}{\hat{\beta}}(R;q)$'s. Therefore, an estimated negative variance may be possible. This occurrence is extremely rare in our experience. However, the WCR variance formula is very unlikely to produce a negative variance estimate when the number of resamples is large and the sample size is sufficient so that asymptotic theory holds. A large number of resamples, $Q \to \infty$, allows the parameter and variance estimates to stabilize. If a negative variance occurs, the simulation is deleted and the coverage is computed as if the simulation did not cover.

A pre-determined, large number of resamples was performed in each resampled data simulated data set. In practice, a convergence criterion that would limit the number of resamples may be implemented if computer resources are limited. This may be developed more in the future.

The GEE approach (Liang and Zeger, 1986) also was used to analyze the simulated data sets and serve as a comparison with the WCR analysis method. In GEE, the exchangeable correlation structure was chosen for all scenarios. This is to emulate a situation where a data analyst does not know the amount of correlation in a data set, but he/she is aware that the data suggest an exchangeable correlation structure, such as with litter-effects data. One attribute of WCR, which is not found in GEE, is that a working covariance matrix does not have to be specified by the user. The WCR variance formula implicitly accounts for the within-cluster correlation. The comparison of the analysis results of WCR and GEE is valid when the cluster sizes are non-infomative. When cluster sizes are informative, we will show by a simulated scenario that GEE is not a consistent analysis method. Thus, it will not be compared with WCR for informative cluster sizes. Informative cluster will be discussed in Chapter V.

The GEE method outputs two estimates for the variance of the marginal parameter: a robust and a model-based. The robust variance, $\underset{\sim}{V}(r)_{GEE}$, uses both the working covariance matrix (exchangeable) and the empirical covariance matrix obtained from the data. The robust variance formula is consistent even when the working correlation matrix is mis-specified. The GEE robust variance formula is

$$\underset{\sim}{V}(r)_{GEE} = \lim_{I\to\infty} I(\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1} \{\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} cov(\underset{\sim}{Y}_i) \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i\} (\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1} \quad (4.10)$$

where, $\underset{\sim}{D}_i = \frac{\partial \mu_i}{\partial \underset{\sim}{\beta}}$ is a matrix of mean derivatives, $\underset{\sim}{V}_i = \underset{\sim}{A}_i^{1/2} R_i(\underset{\sim}{\alpha}) \underset{\sim}{A}_i^{1/2} / \phi$ is a matrix of weights, $\underset{\sim}{A}_i = diag(g(\mu_{ij}))$ is a diagonal matrix of variances, $R_i(\underset{\sim}{\alpha})$ is a $(n_i \text{x } n_i)$ correlation matrix, and $\underset{\sim}{\alpha}$ $(s \text{ x } 1)$ is a vector of parameters that completely specifies $R_i(\underset{\sim}{\alpha})$ (Liang and Zeger, 1986). The other GEE variance is the model-based, $\underset{\sim}{V}(mb)_{GEE}$, which only uses the working covariance matrix

$$\underset{\sim}{V}(mb)_{GEE} = \lim_{I\to\infty} I(\sum_{i=1}^{I} \underset{\sim}{D}_i^T \underset{\sim}{V}_i^{-1} \underset{\sim}{D}_i)^{-1}. \quad (4.11)$$

Liang and Zeger (1986) advocate using the robust variance formula. Therefore, the GEE robust variance, denoted in the result tables as GEE(r), is displayed, unless there is a large discrepancy between the robust and the model-based variances (GEE(mb)). In most cases encountered in this simulation study, the two variances are close in value.

It is important to distinguish between the WCR marginal parameter and the GEE marginal parameter. If the cluster size is non-informative, then both WCR and GEE estimate the same marginal parameter.

The tables containing the results give the general simulation specifications: the average parameter estimates ($\overline{b}_0 = \frac{\sum_{s=1}^{S} \hat{b}_0(s)}{S}; \overline{b}_1 = \frac{\sum_{s=1}^{S} \hat{b}_1(s)}{S}$, where $s = 1, \ldots, S$ indexes the

simulations), empirical standard errors of the parameter estimates ($ese(\widehat{b}_0) = \sqrt{\frac{\sum_{s=1}^{S}(\widehat{b}_0(s)-\overline{b}_0)^2}{S\cdot(S-1)}}; ese(\widehat{b}_1) = \sqrt{\frac{\sum_{s=1}^{S}(\widehat{b}_1(s)-\overline{b}_1)^2}{S\cdot(S-1)}}$), average standard deviations of the estimates ($\overline{std(\widehat{b}_0)} = \frac{\sum_{s=1}^{S}\sqrt{var(\widehat{b}_0(s))}}{S}; \overline{std(\widehat{b}_1)} = \frac{\sum_{s=1}^{S}\sqrt{var(\widehat{b}_1(s))}}{S}$), average coverage of 95% confidence interval for the exposure parameter, the average power (empirical alpha), a paired power comparison (discordant pairs), and the number of zero celled resampled data sets across the $S = 1000$ simulated data sets. The average coverage for the exposure parameter should be 0.95 and the nominal empirical alpha if the null hypothesis, $b_1 = 0$, is true should be 0.05. The paired power comparison reports the number of discordant pairs in favor of WCR or GEE, where the power counts the number of rejections of a 95% confidence interval covering the true parameter value, b_1. McNemar's test (or an exact test based on the binomial distribution when the sample size is too small for a χ^2 test) can be computed when the number of discordant pairs of power rejections in favor of WCR and GEE is known. This test determines if there is a significant difference in favor of one method over the other among the discordant pairs. For all cases, the number of resamples, sample size, and cluster size were varied to see how these factors affected the parameter and variance estimation of the WCR analysis method.

4.2 Null Exposure Effect: Cluster-specific Exposure

4.2.1 Introduction

The first sets of simulations had a null exposure effect, $b_1 = 0$. This situation provided information on the average coverage of 95% confidence intervals for the exposure parameter and the average empirical alpha level under the null hypothesis, H_0: $b_1 = 0$. The null exposure effect simulations were done for zero ($\rho = 0.0$), low ($\rho = 0.2$) and moderately ($\rho = 0.4$) correlated data. The marginal baseline risk was set

to 0.25 for the independent data; and the marginal baseline risks for the low and moderately correlated data were 0.25 and 0.6667, respectively.

For the zero correlation case, the data are generated basically from $binomial(0.25, n_i)$ variates; the low correlation data are generated from $beta(1, 3)$ variates; and the moderate correlation data are generated from $beta(1, 0.5)$ variates. The values of the marginal intercept parameters are as follows:

$$\rho = 0.0: \qquad b_0 = ln\left(\frac{\mu}{1-\mu}\right) = ln\left(\frac{0.25}{0.75}\right) = ln(0.\overline{3}) = -1.0986$$

$$\rho = 0.2: \qquad b_0 = ln\left(\frac{\mu}{1-\mu}\right) = ln\left(\frac{0.25}{0.75}\right) = ln(0.\overline{3}) = -1.0986$$

$$\rho = 0.4: \qquad b_0 = ln\left(\frac{\mu}{1-\mu}\right) = ln\left(\frac{0.67}{0.33}\right) = ln(2) = 0.6931.$$

Recall that the exposure parameter, b_1, was set to 0 for all three levels of within-cluster correlation. Table 4.1. lists the simulation scenarios and indicates the table where the results can be found.

Table 4.1. Null Exposure Simulation Specifications

Table	ρ	Q	I	n_i	b_0	b_1
4.2	0.0	2000 – 20,000	250	2	– 1.0986	0.0
4.3	0.0	2000 – 50,000	250	8	– 1.0986	0.0
4.4	0.2	10,000	250	2; 8	– 1.0986	0.0
4.5	0.4	10,000	250	2; 8	0.6931	0.0
4.6	0.0; 0.2	10,000	250	≤ 8	– 1.0986	0.0
4.6	0.4	10,000	250	≤ 8	0.6931	0.0
4.7	0.0	10,000	250	≤ 8	– 1.0986	0.0

4.2.2 Equal Cluster Sizes

The tables that follow display the results for the null exposure effect case with equal sized clusters. The cluster sizes of $n_i = 2$ and 8, represent a range of cluster sizes that are often found in practice. A cluster size of two is commonly found in ophthalmology studies and larger cluster sizes, such as $n_i = 8$, is common in litter-effects data. It is important to challenge WCR by altering the simulation factors, such as number of resamples, cluster size, sample size, and amount of within-cluster correlation to evaluate the performance of WCR under a range of realistic scenarios. The simulation results for the null exposure effect with equal sized clusters are given in Tables 4.2-4.5.

Table 4.2. Null Exposure Effect, $n_i = 2$, $\rho = 0.0$

	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000
Q	2000		20,000	
I	250	250	250	250
n_i	2	2	2	2
$\bar{b}_0$	−1.1077	−1.1023	−1.1080	−1.1024
$ese(\widehat{b}_0)$	0.0047	0.0047	0.0047	0.0047
$\bar{b}_1$	−0.0008	−0.0007	−0.0064	−0.0064
$ese(\widehat{b}_1)$	0.0065	0.0065	0.0067	0.0066
$\overline{std(\widehat{b}_0)}$	0.1465	0.1467	0.1458	0.1460
$\overline{std(\widehat{b}_1)}$	0.2069	0.2073	0.2071	0.2075
coverage	0.957	0.959	0.941	0.942
empirical alpha	0.043	0.041	0.059	0.058
discordant pairs	3	1	1	0

$b_0 = -1.0986; b_1 = 0.0000.$

Table 4.3. Null Exposure Effect, $n_i = 8$, $\rho = 0.0$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	2000		20,000		50,000	
I	250	250	250	250	400	400
n_i	8	8	8	8	8	8
$\overline{\hat{b}}_0$	−1.1099	−1.1003	−1.1107	−1.1010	−1.1044	−1.0985
$ese(\hat{b}_0)$	0.0023	0.0023	0.0024	0.0024	0.0018	0.0018
$\overline{\hat{b}}_1$	0.0021	0.0022	0.0031	0.0031	0.0025	0.0024
$ese(\hat{b}_1)$	0.0034	0.0033	0.0034	0.0034	0.0026	0.0026
$\overline{std(\hat{b}_0)}$	0.06963	0.0728	0.0693	0.0727	0.0564	0.0579
$\overline{std(\hat{b}_1)}$	0.0986	0.1032	0.0985	0.1031	0.0796	0.0817
coverage	0.919†	0.945	0.927†	0.941	0.939	0.958
empirical alpha	0.081	0.050	0.073	0.059	0.061	0.042
discordant pairs	30*	4	14*	0	19*	0

$b_0 = -1.0986; b_1 = 0.0000;$

† indicates more than 2 standard errors from nominal value, 0.95;

*indicates a significant McNemar's test at 0.05.

Table 4.4. Null Exposure Effect, $\rho = 0.2$

	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000
Q	10,000		10,000	
I	250	250	250	250
n_i	2	2	8	8
$\overline{\hat{b}}_0$	−1.1096	−1.1051	−1.1082	−1.1004
$ese(\hat{b}_0)$	0.0052	0.0051	0.0036	0.0035
$\overline{\hat{b}}_1$	0.0055	0.0054	−0.0046	−0.0045
$ese(\hat{b}_1)$	0.0074	0.0073	0.0052	0.0051
$\overline{std(\hat{b}_0)}$	0.1606	0.1605	0.1120	0.1130
$\overline{std(\hat{b}_1)}$	0.2277	0.2275	0.1590	0.1603
coverage	0.946	0.948	0.940	0.948
empirical alpha	0.054	0.052	0.060	0.052
discordant pairs	2	0	8*	0

$b_0 = -1.0986; b_1 = 0.0000;$

*indicates significant exact binomial test at 0.05.

Table 4.5. Null Exposure Effect, $\rho = 0.4$

	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000
Q	10,000		10,000	
I	250	250	250	250
n_i	2	2	8	8
$\overline{b}_0$	0.7052	0.7033	0.7028	0.6994
$ese(\widehat{b}_0)$	0.0050	0.0050	0.0042	0.0041
$\overline{b}_1$	−0.0068	−0.0068	−0.0015	−0.0014
$ese(\widehat{b}_1)$	0.0071	0.0071	0.0059	0.0058
$std(\widehat{b}_0)$	0.1602	0.1600	0.1306	0.1307
$std(\widehat{b}_1)$	0.2262	0.2259	0.1850	0.1851
coverage	0.952	0.950	0.954	0.954
empirical alpha	0.048	0.050	0.046	0.046
discordant pairs	0	2	1	1

$b_0 = -1.0986; b_1 = 0.0000.$

The null exposure parameter estimates from WCR and GEE are within 2 empirical standard errors $(ese(\widehat{b}_1))$ for all of these scenarios. However, there does appear to be slight bias in all the WCR intercept parameters except Table 4.2 and in the GEE intercept parameters in Tables 4.3 ($Q = 2000$) and Table 4.5 ($n_i = 2$) $(\overline{b}_0 \pm 2 \cdot ese(\widehat{b}_0))$. The WCR variance estimates are smaller than the GEE robust variance estimates for $n_i = 8$, and apparently too small: Table 4.3 reveals poor coverage for WCR. Not only is the coverage significantly different from the nominal value of 0.95, but the WCR coverage is significantly different from the GEE coverage as can be seen by the significant McNemar's tests in Table 4.3. The simulation using only $Q = 2000$ produced an exceptionally poor coverage for WCR in Table 4.3. For $n_i = 2$ in Table 4.2, it appears that increasing the number of resamples from $Q = 2000$ to $Q = 20,000$ does not change the analysis results. However, in Table 4.3 for $n_i = 8$, increasing the number of resamples dramatically improved the coverage. This is in part due to a large number of unstabilized variances that resulted in a much wider range of variance values in WCR than in GEE for $Q = 2000$. It is predominately the extreme variances that reduce the coverage in WCR. It is interesting to point out that GEE also produced slightly under-

nominal average coverages in Table 4.3, although these values were not statistically different from 0.95.

More resamples are evidently required as the within-cluster correlation goes to zero, $\rho \rightarrow 0$, or as I increases. When the correlation in a data set decreases, the resampled data sets are less correlated, less alike. Therefore, to capture an appropriate representation of the resampling distribution, more resamples are needed to obtain stable parameter and variance estimates. As the within-cluster correlation increases, the resampled data sets are more similar, and fewer resamples are necessary to approximate the resampling distribution.

However, since WCR was developed for clustered data with positive within-cluster correlation, the results in Table 4.4 and Table 4.5 provide a more realistic view of the behavior of the WCR method. The average coverage for WCR and GEE are not statistically different from 0.95, and as the within-cluster correlation increases, the similarity between WCR and GEE increases. The two methods produced almost identical results in Table 4.4 for $n_i = 2$ and in Table 4.5. The paired power comparison in Table 4.4 showed a significant difference for $n_i = 8$ since all of the discordant pairs had WCR rejecting and GEE not.

4.2.3 Unequal Cluster Sizes

Unequal cluster sizes were generated by using a random binomial generator in C to determine the cluster size, n_i. The zero sized clusters were deleted so the cluster size is actually generated from a truncated, $binomial(0.5, 8)$. The results for the null exposure effect simulations with unequal cluster sizes are given in Tables 4.6.

Table 4.6. Null Exposure Effect, $n_i \leq 8$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)	GEE(mb)
S	1000	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000		
I	250	250	250	250	250	250	250
n_i	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8
ρ	0.0	0.0	0.2	0.2	0.4	0.4	0.4
$\bar{b}_0$	−1.1066	−1.1007	−1.1051	−1.0988	0.6969	0.6945	0.6945
$ese(\widehat{b}_0)$	0.0036	0.0033	0.0044	0.0043	0.0048	0.0047	0.0047
$\bar{b}_1$	−0.0068	−0.0042	−0.0056	−0.0026	−0.0033	−0.0021	−0.0021
$ese(\widehat{b}_1)$	0.0051	0.0046	0.0062	0.0060	0.0067	0.0067	0.0067
$std(\widehat{b}_0)$	0.1112	0.1028	0.1363	0.1343	0.1445	0.1466	0.1416
$std(\widehat{b}_1)$	0.1575	0.1459	0.1929	0.1900	0.2047	0.2079	0.2005
coverage	0.945	0.952	0.945	0.942	0.941	0.946	0.937
empirical alpha	0.055	0.048	0.055	0.058	0.059	0.054	0.063
discordant pairs	20	13	19	22	22/15	17	19

$\rho = 0.0; 0.2: \; b_0 = -1.0986; b_1 = 0.0000;$
$\rho = 0.4: \; b_0 = 0.6931; b_1 = 0.0000.$

The WCR and GEE estimation results for all values of ρ in Table 4.6 are without significant bias, except for $\rho = 0.0$ where the WCR intercept parameter is slightly biased downward. All of the simulations have the appropriate average coverage and empirical alpha. The GEE model-based variance simulation results are given for $\rho = 0.4$ because they are quite different from the GEE robust variance simulations. The model-based variances are smaller than both the WCR and GEE robust variance estimates. It could be argued that the model-based variance is an appropriate variance since the data are beta-binomial, which is known to have an exchangeable correlation structure, but this would typically not be known by the investigator. Although all the average estimates, average coverages, and average empirical alphas are in the correct ranges, the GEE model-based variance may be too small.

4.2.4 Unequal Cluster Sizes from a Mixture of Binomials

The next set of scenarios generated the cluster sizes as a mixture of two binomials. A random one-fourth of the clusters have their cluster size generated from a truncated, $binomial(0.75, 8)$, and the remaining clusters have their cluster size generated from a truncated, $binomial(0.25, 8)$. Therefore, the majority of the clusters are small in size with only a few large clusters. These simulation results are displayed in Table 4.7.

Table 4.7. Null Exposure Effect with Cluster Size from a Mixture, $\rho = 0.0$

	WCR	GEE(r)
S	1000	1000
Q	10,000	
I	250	250
n_i	≤ 8	≤ 8
$\overline{b}_0$	-1.0992	-1.0978
$ese(\widehat{b}_0)$	0.0045	0.0035
$\overline{b}_1$	-0.0119	-0.0093
$ese(\widehat{b}_1)$	0.0063	0.0050
$std(\widehat{b}_0)$	0.1414	0.1143
$std(\widehat{b}_1)$	0.2002	0.1622
coverage	0.955	0.959
empirical alpha	0.045	0.041
discordant pairs	45	41

$b_0 = -1.0986; b_1 = 0.0000.$

It is interesting to note that generating the cluster size from two binomials should not affect the analysis. The cluster size is unequal and not related to the baseline risk thus non-informative. Although, the average parameter estimates and average coverage are unbiased, the variances from the WCR simulations are much larger than from the GEE simulations. Comparing the variance components from the simulations with unequal clusters in Table 4.6 with those in Table 4.7 showed that the first component of the WCR variance is virtually identical in the two scenarios. However, the second component of the variance was smaller in Table 4.7; therefore, a larger variance was found in Table 4.7

than Table 4.6. There are more clusters in scenario of Table 4.7 that are of size 1, so there is considerable overlap in the resamples. Thus, the second component of the WCR variance in Table 4.7 is likely to be smaller than its counterpart in Table 4.6, so the WCR variance in Table 4.7 is larger than in Table 4.6. This makes sense intuitively since the average cluster size is $\overline{n}_i = 3.1669$ in Table 4.7, which is smaller than the average cluster size, $\overline{n}_i = 4.1113$, in Table 4.6.

4.3 Positive Exposure Effect: Cluster-specific Exposure

4.3.1 Introduction

The case of positive exposure effects with cluster-specific exposures is a more interesting and realistic scenario for the simulations. The positive exposure effect is achieved by adding a term on the logit scale, which increases (decreases) the risk for the outcome of interest when the exposure is present. The cluster-specific model can be written as in (4.2), however, the within-cluster parameter, α is no longer set to 0

Cluster-specific Model:

$$ln(\frac{E(Y_{ij}|u_i,\mathrm{E}_{ij})}{1-E(Y_{ij}|u_i,\mathrm{E}_{ij})}) = ln(\frac{u_i}{1-u_i}) + \alpha \mathrm{E}_{ij}, \tag{4.12}$$

where u_i is defined as a $beta(a,b)$ variate. The value of α was set so that both the low and moderately correlated data gave the same marginal exposure parameter, $b_1 = 0.4796$. The following are the marginal risks for the exposed clusters

$\rho = 0.2, \alpha = 0.6111$:

$$E(Y_{ij}|\text{exposed}) = \int_0^1 \frac{exp(\pi_i+\alpha)}{1+exp(\pi_i+\alpha)}(\frac{1}{B(1,3)})u_i^0(1-u_i)^2 du_i = 0.35 \tag{4.13}$$

$\rho = 0.4, \alpha = 0.7834$:

$$E(Y_{ij}|\text{exposed}) = \int_0^1 \frac{exp(\pi_i+\alpha)}{1+exp(\pi_i+\alpha)} (\frac{1}{B(1,0.5)}) u_i^0 (1-u_i)^{-0.5} du_i = 0.7638. \qquad (4.14)$$

Recall that the marginal baseline risks are 0.25 for $\rho = 0.0$ and $\rho = 0.2$ and 0.6667 for $\rho = 0.4$. The marginal model and its corresponding parameters can be found from the values of the marginal risk for unexposed and exposed observations.

Marginal Model:

$$\begin{aligned} ln(\frac{E_i[E_{j|i}(Y_{ij}|\mathrm{E}_{ij})]}{1-E_i[E_{j|i}(Y_{ij}|\mathrm{E}_{ij})]}) &= ln(\frac{\mu}{1-\mu}) + \alpha^* \mathrm{E}_{ij} \\ &= b_0 + b_1 E_{ij} \end{aligned} \qquad (4.15)$$

The values of the marginal intercept and marginal exposure parameter are $b_0 = -1.0986$ and $b_1 = 0.4796$, respectively, for $\rho = 0.0$ and $\rho = 0.2$. The marginal intercept and marginal exposure parameter for $\rho = 0.4$ are $b_0 = 0.6931$ and $b_1 = 0.4796$, respectively. These values of the parameters were calculated assuming non-informative cluster sizes. The corresponding odds-ratio is 1.6154.

The traditional interpretation of the odds-ratio is the ratio of odds for exposed observations versus unexposed observations with and without the outcome of interest. The interpretation difference between a WCR parameter and a GEE parameter is subtle. The WCR method is a cluster-based analysis method, whereas GEE is an observation-based analysis method. Therefore, the WCR parameter incorporates the sampling scheme of the method into the interpretation of its parameters. The interpretation of the WCR parameter estimate is the difference in log odds for the outcome of interest between a randomly sampled exposed observation from a randomly selected cluster and a randomly

sampled unexposed observation from a randomly selected cluster. This interpretation of the WCR parameter estimate reflects the one-per-cluster sampling scheme of the procedure. Table 4.8, provides the simulation specifications for the positive exposure scenarios.

Table 4.8. Positive Exposure Simulation Specifications

Table	Covariate	ρ	Q	I	n_i	b_0	b_1
4.9	CS	0.0	$2000 - 20,000$	250	2	-1.0986	0.4796
4.10	CS	0.0	$2000 - 50,000$	$250 - 400$	8	-1.0986	0.4796
4.11	CS	0.2	10,000	$50 - 250$	2	-1.0986	0.4796
4.12	CS	0.2	$10,000 - 40,000$	$50 - 250$	8	-1.0986	0.4796
4.13	CS	0.4	10,000	$50 - 250$	2	0.6931	0.4796
4.14	CS	0.4	10,000	$50 - 250$	8	0.6931	0.4796
4.16	CS	0.0	10,000	250	≤ 8	-1.0986	0.4796
4.17	CS	0.2	10,000	$50 - 250$	≤ 8	-1.0986	0.4796
4.18	CS	0.4	10,000	$50 - 250$	≤ 8	0.6931	0.4796
4.19	CS	0.0	10,000	250	≤ 8	-1.0986	0.4796
4.20	OS	0.2	$10,000 - 50,000$	250	$2; 8; \leq 8$	-1.0986	0.4796
4.21	OS	0.4	$10,000 - 50,000$	250	$2; 8; \leq 8$	0.6931	0.4796

4.3.2 Equal Cluster Sizes

The following scenarios have equal sized clusters, $n_i = 2,\ 8\ \forall i$. All cases vary the number of resamples and sample size. Up to this point, there have not been any simulations with zero cells. However by decreasing the number of clusters in the simulations, zero cells occurred. The total number of zero cells out of the 1000 simulated data sets is given in each table (# zero cells out of $S \cdot Q$) when this occurred. The simulation results for the positive exposure effect with equal sized clusters are displayed in Tables 4.9-4.14.

Table 4.9. Positive Exposure Effect, $n_i = 2$, $\rho = 0.0$

	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000
Q	2000		20,000	
I	250	250	250	250
n_i	2	2	2	2
$\overline{\hat{b}}_0$	−1.1069	−1.1012	−1.1120	−1.1065
$ese(\hat{b}_0)$	0.0046	0.0046	0.0048	0.0047
$\overline{\hat{b}}_1$	0.4785	0.4756	0.4844	0.4817
$ese(\hat{b}_1)$	0.0063	0.0063	0.0062	0.0062
$\overline{std(\hat{b}_0)}$	0.1458	0.1462	0.1462	0.1464
$\overline{std(\hat{b}_1)}$	0.1975	0.1979	0.1979	0.1981
coverage	0.951	0.954	0.954	0.955
power	0.680	0.677	0.701	0.697
discordant pairs	7	4	4	0

$b_0 = -1.0986; b_1 = 0.4796.$

Table 4.10. Positive Exposure Effect, $n_i = 8$, $\rho = 0.0$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	2000		20,000		50,000	
I	250	250	250	250	400	400
n_i	8	8	8	8	8	8
$\overline{\hat{b}}_0$	−1.1110	−1.1013	−1.1073	−1.0976	−1.1057	−1.0997
$ese(\hat{b}_0)$	0.0024	0.0023	0.0023	0.0023	0.0018	0.0018
$\overline{\hat{b}}_1$	0.4866	0.4819	0.4863	04814	0.4839	0.4809
$ese(\hat{b}_1)$	0.0031	0.0031	0.0031	0.0031	0.0025	0.0025
$\overline{std(\hat{b}_0)}$	0.0697	0.0729	0.0696	0.0728	0.0563	0.0577
$\overline{std(\hat{b}_1)}$	0.0940	0.0986	0.0948	0.0985	0.0763	0.0781
coverage	0.936	0.964†	0.935	0.956	0.943	0.948
power	0.998	1.000	0.999	0.999	1.000	1.000
discordant pairs	0	1	0	0	0	0

$b_0 = -1.0986; b_1 = 0.4796;$

† indicates more than 2 standard errors from nominal value, 0.95.

Table 4.11. Positive Exposure Effect, $n_i = 2$, $\rho = 0.2$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000	
I	50	50	100	100	250	250
n_i	2	2	2	2	2	2
$\overline{\hat{b}_0}$	−1.1457	−1.1188	−1.1299	−1.1173	−1.1139	−1.1094
$ese(\hat{b}_0)$	0.0122	0.0118	0.0084	0.0082	0.0052	0.0052
$\overline{\hat{b}_1}$	0.4884	0.4749	0.5117	0.5046	0.4889	0.4865
$ese(\hat{b}_1)$	0.0168	0.0163	0.0113	0.0111	0.0070	0.0070
$\overline{std(\hat{b}_0)}$	0.3640	0.3618	0.2562	0.2561	0.1609	0.1608
$\overline{std(\hat{b}_1)}$	0.4969	0.4944	0.3470	0.3466	0.2185	0.2184
coverage	0.930†	0.934†	0.948	0.952	0.944	0.945
power	0.170	0.157	0.310	0.307	0.627	0.625
discordant pairs	14*	1	3	0	3	1
# zero cells out of $S \cdot Q$	8729		2		0	

$b_0 = -1.0986; b_1 = 0.4796;$

† indicates more than 2 standard errors from nominal value, 0.95;

*indicates significant McNemar's test at 0.05.

Table 4.12. Positive Exposure Effect, $n_i = 8$, $\rho = 0.2$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000	1000	1000
Q	10,000		40,000		10,000		10,000	
I	50	50	50	50	100	100	250	250
n_i	8	8	8	8	8	8	8	8
$\overline{\hat{b}_0}$	−1.1642	−1.1191	−1.1611	−1.1166	−1.1356	−1.1144	−1.1102	−1.1023
$ese(\hat{b}_0)$	0.0086	0.0081	0.0088	0.0084	0.0061	0.0060	0.0036	0.0035
$\overline{\hat{b}_1}$	0.5148	0.4917	0.5041	0.4817	0.5028	0.4914	0.4833	0.4791
$ese(\hat{b}_1)$	0.0119	0.0113	0.0118	0.0112	0.0082	0.0080	0.0049	0.0048
$\overline{std(\hat{b}_0)}$	0.2379	0.2470	0.2401	0.2489	0.1737	0.1784	0.1122	0.1131
$\overline{std(\hat{b}_1)}$	0.3343	0.3453	0.3371	0.3479	0.2419	0.2465	0.1551	0.1559
coverage	0.905†	0.940	0.926†	0.940	0.920†	0.940	0.950	0.952
power	0.360	0.312	0.322	0.279	0.543	0.515	0.873	0.872
discordant pairs	49*	1	43*	0	28*	0	2	1
# zero cells out of $S \cdot Q$	11,825		48,697		11		0	

$b_0 = -1.0986; b_1 = 0.4796;$

† indicates more than 2 standard errors from nominal value, 0.95;

*indicates significant McNemar's test at 0.05.

Table 4.13. Positive Exposure Effect, $n_i = 2$, $\rho = 0.4$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000	
I	50	50	100	100	250	250
n_i	2	2	2	2	2	2
$\overline{\hat{b}}_0$	0.7395	0.7274	0.7019	0.6970	0.6966	0.6947
$ese(\hat{b}_0)$	0.0126	0.0123	0.0082	0.0082	0.0053	0.0052
$\overline{\hat{b}}_1$	0.5150	0.5022	0.4955	0.4898	0.4858	0.4838
$ese(\hat{b}_1)$	0.0179	0.0174	0.0127	0.0125	0.0076	0.0075
$\overline{std(\hat{b}_0)}$	0.3638	0.3619	0.2531	0.2526	0.1594	0.1592
$\overline{std(\hat{b}_1)}$	0.5464	0.5413	0.3791	0.3781	0.2370	0.2368
coverage	0.943	0.945	0.939	0.943	0.939	0.940
power	0.159	0.150	0.250	0.246	0.514	0.513
discordant pairs	11*	2	4	0	1	0
# zero cells out of $S \cdot Q$	7829		25		0	

$b_0 = 0.6931; b_1 = 0.4796;$
*indicates significant McNemar's test at 0.05.

Table 4.14. Positive Exposure Effect, $n_i = 8$, $\rho = 0.4$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000	
I	50	50	100	100	250	250
n_i	8	8	8	8	8	8
$\overline{\hat{b}}_0$	0.7368	0.7168	0.6983	0.6897	0.6945	0.6912
$ese(\hat{b}_0)$	0.0099	0.0096	0.0070	0.0069	0.0043	0.0043
$\overline{\hat{b}}_1$	0.5144	0.4937	0.5153	0.5050	0.4864	0.4829
$ese(\hat{b}_1)$	0.0146	0.0140	0.0100	0.0098	0.0064	0.0064
$\overline{std(\hat{b}_0)}$	0.2913	0.2927	0.2063	0.2065	0.1310	0.1311
$\overline{std(\hat{b}_1)}$	0.4351	0.4328	0.3049	0.3056	0.1928	0.1929
coverage	0.937	0.945	0.947	0.953	0.931†	0.933†
power	0.237	0.221	0.405	0.389	0.715	0.714
discordant pairs	21*	5	16*	0	3	2
# zero cells out of $S \cdot Q$	19,193		14		0	

$b_0 = 0.6931; b_1 = 0.4796;$
† indicates more than 2 standard errors from nominal value, 0.95;
*indicates significant McNemar's test at 0.05.

There is some slight bias associated with the parameter estimates in the WCR and GEE simulation results in Tables 4.9-4.14. Tables 4.9 and 4.10 are the only scenarios where there is only significant bias in the WCR parameter estimates, otherwise, the bias for WCR and GEE parameter estimates tend to go together. The positive exposure cases are very close in variance estimates, average coverage and average power as the number of resamples, sample size, or correlation increases. WCR does have slightly better power than GEE for $n_i = 2$, which the paired power comparison showed that there is a significant difference in the direction of the discordant pairs favoring WCR. In Table 4.10, the GEE coverage is too large, which suggests that the robust variance for this case is too big. When $Q = 2000$ in Table 4.10, one simulation produced a negative variance in the WCR analyses. This simulation was not included in the calculation of the parameter and variance estimates, but was included for the coverage calculation as not covering. Although each of the three different scenarios portrayed in Table 4.10 used different data sets, increasing the number of resamples stabilized the variance estimates and the occurrence of a negative variance did not subsequently occur. In Table 4.12, the simulations with small sample size, $I = 50, 100$, produced poor results for WCR. However, by increasing the number of resamples from $Q = 10,000$ to $50,000$ or increasing the number of clusters to $I = 250$, WCR performed much better with regards to average coverage across the resamples. WCR does have better average power in many of these scenarios (Tables 4.9, 4.11-4.14). The paired power comparisons in these tables found significant differences in the direction of the discordant pairs favoring WCR.

4.3.3 Unequal Cluster Sizes

The two methods under consideration, WCR and GEE, were compared when the cluster sizes were unequal, as in Section 4.2.3, except now with positive exposure effects. The independent data are generated from $binomial(0.25, n_i)$ variates so the baseline risk is 0.25 and increases to 0.35 when a cluster-specific component is added on the logit

scale. The low correlation data are generated from $beta(1,3)$ variates so the baseline risk is 0.25 and also increases to 0.35. The moderately correlated data are generated from $beta(1,0.5)$ variates so the baseline risk is 0.6667 and increases to 0.7638. Table 4.15 gives details on the value of the cluster-specific component and the marginal risks and the simulation results for a positive exposure effect with unequal cluster sizes are displayed in Tables 4.16-4.18.

Table 4.15. Marginal Risks for Unexposed and Exposed Groups

	Within-cluster Component (α)	WCR Marginal Risk-Unexposed Group	WCR Marginal Risk-Exposed Group
$\rho = 0.0$	0.6111	0.25	0.35
$\rho = 0.2$	0.6111	0.25	0.35
$\rho = 0.4$	0.7834	0.6667	0.7638

Table 4.16. Positive Exposure Effect, $n_i \leq 8$, $\rho = 0.0$

	WCR	GEE(r)
S	1000	1000
Q	10,000	
I	250	250
n_i	≤ 8	≤ 8
$\overline{\hat{b}_0}$	-1.1144	-1.1042
$ese(\hat{b}_0)$	0.0037	0.0033
$\overline{\hat{b}_1}$	0.4935	0.4870
$ese(\hat{b}_1)$	0.0047	0.0043
$\overline{std(\hat{b}_0)}$	0.1112	0.1036
$\overline{std(\hat{b}_1)}$	0.1503	0.1389
coverage	0.953	0.959
power	0.906	0.948
discordant pairs	10	52*

$b_0 = -1.0986; b_1 = 0.4796;$
*indicates significant McNemar's test at 0.05.

Table 4.17. Positive Exposure Effect, $n_i \leq 8$, $\rho = 0.2$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000	
I	50	50	100	100	250	250
n_i	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8
$\overline{b}_0$	−1.1642	−1.1171	−1.1256	−1.1122	−1.1105	−1.1058
$ese(\widehat{b}_0)$	0.0101	0.0101	0.0073	0.0071	0.0045	0.0044
$\overline{b}_1$	0.5158	0.4849	0.5018	0.4944	0.4807	0.4776
$ese(\widehat{b}_1)$	0.0147	0.0139	0.0093	0.0091	0.0059	0.0059
$\overline{std(\widehat{b}_0)}$	0.3055	0.2988	0.2140	0.2117	0.1362	0.1342
$\overline{std(\widehat{b}_1)}$	0.4193	0.4125	0.2942	0.2907	0.1865	0.1843
coverage	0.933†	0.940	0.947	0.954	0.949	0.950
power	0.259	0.239	0.387	0.389	0.732	0.741
discordant pairs	62*	42	60	62	49	58
# zero cells out of $S \cdot Q$	13,847		12		0	

$b_0 = -1.0986; b_1 = 0.4796;$

*indicates significant McNemar's test at 0.05.

Table 4.18. Positive Exposure Effect, $n_i \leq 8$, $\rho = 0.4$

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		10,000		10,000	
I	50	50	100	100	250	250
n_i	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8
$\overline{b}_0$	0.7244	0.7113	0.7061	0.7057	0.6938	0.6922
$ese(\widehat{b}_0)$	0.0110	0.0111	0.0074	0.0074	0.0043	0.0044
$\overline{b}_1$	0.5201	0.5047	0.5082	0.4927	0.4962	0.4936
$ese(\widehat{b}_1)$	0.0163	0.0161	0.0109	0.0107	0.0066	0.0067
$\overline{std(\widehat{b}_0))}$	0.3270	0.3289	0.2295	0.2323	0.1445	0.1467
$\overline{std(\widehat{b}_1)}$	0.4906	0.4850	0.3409	0.3438	0.2140	0.2167
coverage	0.946	0.943	0.956	0.954	0.963‡	0.959
power	0.193	0.185	0.325	0.302	0.636	0.627
discordant pairs	48	40	68*	45	64	55
# zero cells out of $S \cdot Q$	17,324		14		0	

$b_0 = 0.6931; b_1 = 0.4796;$

†indicates more than 2 standard errors from nominal value, 0.95;

*indicates significant McNemar's test at 0.05.

Many of the WCR parameter estimates in these scenarios are slightly biased, whereas the corresponding GEE parameter estimates are not. The WCR variances are larger for $\rho = 0.0$ and 0.2 than those for GEE. The average coverages are within two standard errors of 0.95, except for the WCR simulation case of $I = 50$ in Table 4.17. Although no conclusion can be drawn regarding power, WCR has the advantage in Tables 4.17 and 4.18, whereas GEE has the advantage in Table 4.16 as seen by the paired power comparisons (McNemar's test on the discordant pairs).

4.3.4 Unequal Cluster Size from a Mixture of Binomials

The next case is the positive exposure equivalent of Section 4.2.4. The simulation results for this are displayed in Table 4.19.

Table 4.19. Positive Exposure Effect with Cluster Size from a Mixture, $\rho = 0.0$

	WCR	GEE(r)
S	1000	1000
Q	10,000	
I	250	250
n_i	≤ 8	≤ 8
$\overline{b}_0$	-1.1108	-1.1017
$ese(\widehat{b}_0)$	0.0046	0.0038
$\overline{b}_1$	0.4841	0.4811
$ese(\widehat{b}_1)$	0.0061	0.0050
$\overline{std(\widehat{b}_0)}$	0.1417	0.1148
$\overline{std(\widehat{b}_1)}$	0.1914	0.1560
coverage	0.954	0.945
power	0.717	0.874
discordant pairs	14	171*

$b_0 = -1.0986; b_1 = 0.4796;$

*indicates a significant McNemar's test at 0.05.

For the same reasons as given in Section 4.2.4, the WCR variance estimates again are much larger than the corresponding GEE robust variance estimates. The GEE method also has much greater average power, which can be seen by the large number of

discordant pairs in favor of GEE. However, the parameter estimates and coverages for both methods are close and without bias $(\overline{b}_0 \pm 2 \cdot ese(\widehat{b}_0); \overline{b}_1 \pm 2 \cdot ese(\widehat{b}_1))$.

4.4 Positive Exposure Effect: Observation-specific Exposure

The same type of simulation study was performed when there were observation-specific, positive exposures for positively correlated data. The exposure status is randomly assigned to the observations (with probability 0.5) instead of the clusters as in the previous simulations. The marginal parameter are the same as with cluster-specific covariates. The following two tables, Table 4.20 and Table 4.21, display the simulation results for observation-specific, positive exposure effect data analyzed by both WCR and GEE exchangeable.

Table 4.20. Positive Exposure Effect, $\rho = 0.2$ (OS)

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		50,000		50,000	
I	250	250	250	250	250	250
n_i	2	2	≤ 8	≤ 8	8	8
$\overline{b}_0$	-1.1089	-1.1041	-1.1106	-1.1031	-1.1095	-1.1007
$ese(\widehat{b}_0)$	0.0050	0.0049	0.0041	0.0039	0.0031	0.0029
$\overline{b}_1$	0.4835	0.4816	0.4829	0.4807	0.4858	0.4807
$ese(\widehat{b}_1)$	0.0064	0.0063	0.0049	0.0044	0.0032	0.0029
$\overline{std(\widehat{b}_0)}$	0.1537	0.1522	0.1244	0.1168	0.0936	0.0936
$\overline{std(\widehat{b}_1)}$	0.1979	0.1930	0.1511	0.1327	0.0950	0.0916
coverage	0.950	0.939	0.944	0.942	$0.928^{\dagger}$	0.954
power	0.704	0.712	0.882	0.944	0.999	0.999
discordant pairs	28	36	10	72^{*}	0	0

$b_0 = -1.0986; b_1 = 0.4796;$

† indicates more than 2 standard errors from nominal value, 0.95;

*indicates significant McNemar's test at 0.05.

Table 4.21. Positive Exposure Effect, $\rho = 0.4$ (OS)

	WCR	GEE(r)	WCR	GEE(r)	WCR	GEE(r)
S	1000	1000	1000	1000	1000	1000
Q	10,000		50,000		50,000	
I	250	250	250	250	250	250
n_i	2	2	≤ 8	≤ 8	8	8
$\overline{b}_0$	0.6984	0.6965	0.6995	0.6982	0.6993	0.6955
$ese(\widehat{b}_0)$	0.0047	0.0045	0.0041	0.0038	0.0034	0.0033
$\overline{b}_1$	0.4897	0.4852	0.4877	0.4808	0.4817	0.4759
$ese(\widehat{b}_1)$	0.0065	0.0060	0.0049	0.0039	0.0032	0.0027
$\overline{std(\widehat{b}_0)}$	0.1473	0.1427	0.1256	0.1166	0.1032	0.1004
$\overline{std(\widehat{b}_1)}$	0.2012	0.1863	0.1542	0.1239	0.0976	0.0847
coverage	0.944	0.951	0.949	0.948	0.932†	0.939
power	0.674	0.746	0.882	0.974	0.997	1.0
discordant pairs	34	106*	3	95*	0	3

$b_0 = 0.6931; b_1 = 0.4796;$
† indicates more than 2 standard errors from nominal value, 0.95;
*indicates significant McNemar's test at 0.05.

As previously shown, the two analysis methods perform very similarly in this simulation study. Although the WCR intercept parameter estimates tend to be slightly biased, the WCR and GEE exposure parameter estimates are within 2 standard errors of their target values. The power for GEE is slightly higher and the variance estimates were smaller, which was supported by the paired power comparisons. A large number of resamples are required for the observation-specific case since the number of distinct resampled-data sets has increased.

4.5 Q-Q Plots

Q-Q plots were done for each WCR simulation of Chapter IV. A Q-Q plot is the 2000 standardized parameter estimates from a WCR simulation plotted against the quantiles of the standard normal distribution. If the data points fall on a 45° line, then the data suggest that the distribution of the parameter is standard normal. The Q-Q plots were visual evidence of the asymptotic normality of a standardized WCR estimator. A Q-

Q plot was done for both the intercept and the exposure parameter for all of the simulations in Chapter IV. The Q-Q plot is titled according to which table it corresponds with additional information in a sub-title when more than one simulation is presented in a table in Chapter IV. The Q-Q plots corresponding to the simulations of Chapter IV can be found in Appendix 1. As can be seen by the Q-Q plots in Appendix 1, the data are very suggestive that the parameter estimates are asymptotically normal.

4.6 Summary

Throughout the simulations, WCR, in general, performed very well when compared with GEE. WCR tends to produce slightly more biased parameter estimates than GEE, especially the intercept parameter estimates in these scenarios. We believe WCR takes longer to reach asymptotics than GEE, thus causing more bias with smaller sample sizes, fewer resamples, and less within-cluster correlation.

The two analysis methods are attempting to estimate the same parameters, but by different means. For independent data, WCR does not perform as well as GEE, especially with larger cluster sizes. As the within-cluster correlation increases, the WCR method does much better in estimating the correct variance. The WCR power when there are unequal cluster sizes tends to be lower than GEE.

One advantage of WCR is that it does not require specifying a working correlation matrix, which often is a guess. The WCR variance formula implicitly accounts for the within-cluster correlation, no matter what form it may take. The other advantage of WCR is the weighting of the clusters. WCR is a cluster-based analysis method which has a different parameter interpretation when there are informative cluster sizes. The WCR sampling scheme weights each cluster equally so that larger clusters are not given special status in an analysis. The definition of informative cluster sizes, a scenario where GEE is invalid, simulation results, and a data analysis are in Chapter V.

CHAPTER V

INFORMATIVE CLUSTER SIZES

5.1 Introduction

In many research situations, a study design produces data which is considered to be "clustered". A cluster is a group of dependent outcomes, which may be specific to an individual, an animal, or a plot of land. The outcomes nested within a cluster will be referred to as observations. Examples might include whether or not a particular pregnancy of a woman resulted in a live-birth; or whether or not a pup from a litter has a birth-defect. Although distinct clusters are often assumed independent, the dependency among the observations within a cluster is referred to as within-cluster correlation and should not be ignored. By analyzing positively correlated data with a method that ignores the within-cluster correlation, such as ordinary logistic regression, the variances tend to be under-estimated, which may lead to inflated Type I error rates.

This chapter discusses an additional issue sometimes found in clustered data which will be referred to as "informative cluster sizes". In Section 5.2, the term *informative cluster size* is defined. Section 5.3 describes how informative cluster sizes affect the traditional marginal parameter, as well as introducing Within Cluster Resampling (WCR) as a valid method for analyzing informative cluster size data. Simulation results are presented in Section 5.4. An analysis of an toxicological data set is given in Section 5.6. Concluding remarks regarding the issue of informative cluster sizes and the appropriateness of analysis methods are presented in Section 5.7.

5.2 Definition

In addition to within-cluster correlation in clustered data, the existence of a relationship between the size of the cluster and the baseline risk for the outcome of interest may be of some concern in an analysis. This relationship will be referred to as "informative cluster sizes". Most analysis methods implicitly assume that cluster sizes are not informative. However, WCR is a valid analysis method under informative cluster sizes, as well as non-informative cluster sizes. This feature of WCR makes it an attractive analysis method, since the existence of informative cluster sizes is often not known.

An example of informative cluster sizes can be found in litter-effects studies. Suppose hypothetically that the mean risk in a population of litters is 0.5. Assume there is an equal number of low risk dams, 0.1, and high risk dams, 0.9. Further if we suppose the low risk dams have litters of size 1 and the high risk dams have litters of size 10, then it is clear that under a traditional marginal model the parameter is not consistent for the dam-based mean risk of 0.5. However, the WCR parameter is consistent due to the equal weighting of the dams. A less extreme situation can realistically arise since some dams naturally produce large litters and other dams more naturally are inclined to produce small litters. If the dams who produce larger litters are more prone to have pups with the outcome of interest and the smaller litters are less likely to have pups with the outcome of interest, then there is a problem due to informative cluster sizes. The problem is evident in human studies as well. With pregnancy history data, if women of child-bearing age are studied and the outcome of interest is spontaneous abortion, then women who had many pregnancies may have a higher proportion of spontaneous abortion, than the women who had fewer pregnancies.

Informative cluster sizes invalidate traditional marginal analysis approaches primarily because the weighting of the observations in the analysis produces inconsistent estimates. For this reason, traditional marginal analysis methods are not valid for analyzing informative cluster size data. We will focus on the WCR method as a valid

alternative and demonstrate the inconsistent results GEE gives under an informative cluster size scenario. This strategy will be discussed in greater detail in Sections 5.3 and 5.4.

5.3 Marginal Parameter

We will differentiate the difference between an observation-based and a cluster-based parameter. In traditional marginal analyses, the marginal parameter is observation-based, meaning that the unit of analysis is the observation. For example, with the pregnancy history data, the pregnancies are the units of analysis. The interpretation of the marginal parameter is the population-averaged difference on the link-transformed scale corresponding to a unit difference in the covariate. The effect of the covariates is averaged across the clusters and observations. In a logistic model, the difference in log odds between the two exposure groups is computed. Looking at an example data set (Paul, 1982) in the toxicology setting, it is easy to define the traditional marginal parameter. The Paul data set consists of banded Dutch rabbits that were observed for skeletal and visceral abnormalities. There were 84 litters with the litter size ranging from 1 to 13 rabbit pups. There were a control and three treatment groups (low, medium, and high) based on prenatal administration of a compound that was thought to cause abnormalities in pups.

The traditional marginal risk would be the probability that a *randomly selected pup* is abnormal. Using the low treatment group from the data set in Table 5.8, this risk is computed as

$$\widehat{Pr}_{(\text{O-B})}(\text{abnormal}) = \frac{\text{total number of abnormal pups}}{\text{total number of pups}}$$

$$= \frac{(0+1+1+0+2+0+1+0+1+0+0+3+0+0+1+5+0+0+3)}{(5+11+7+9+12+8+6+7+6+4+6+9+6+7+5+9+1+6+9)} = 0.135$$

where (O-B) denotes observation-based. Notice that the larger litters are given greater weight than the smaller litters. When informative cluster sizes are present this weighting scheme has an effect on estimation (even asymptotically) since the higher risk clusters are of either greater or lesser size than the lower risk clusters. WCR avoids this by giving equal weight to all clusters.

A cluster-based, marginal risk can be computed where instead of giving each observation equal weight, each cluster is equally weighted. This risk can be estimated as follows using the low treatment group in the Paul data set in Table 5.8

$$\widehat{Pr}_{(\text{C-B})}(\text{abnormal}) = \frac{\text{sum of litter-specific abnormality } \hat{p}\text{'s}}{\text{number of litters}}$$

$$= \frac{(\frac{0}{5}+\frac{1}{11}+\frac{1}{7}+\frac{0}{9}+\frac{2}{12}+\frac{0}{8}+\frac{1}{6}+\frac{0}{7}+\frac{1}{6}+\frac{0}{4}+\frac{0}{6}+\frac{3}{9}+\frac{0}{6}+\frac{0}{7}+\frac{1}{5}+\frac{5}{9}+\frac{0}{1}+\frac{0}{6}+\frac{3}{9})}{19} = 0.113$$

where (C-B) denotes cluster-based. This cluster-based, marginal risk in WCR is the probability that a *randomly selected pup from a randomly selected litter* is abnormal. Note that the difference is in the weighting of the clusters. The cluster-based approach *equally weights* each litter, whereas, in an observation-based approach *larger* litters have *greater weight* than small litters.

The literature on clustered binary data analysis methods does not recognize this distinction in marginal parameters and simply implicitly assumes there are non-informative cluster sizes. One such method is the generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986). WCR provides an analysis method which is valid for data with informative cluster sizes. Also, WCR estimates observation-based marginal parameters in the absence of informative cluster sizes and estimates cluster-based marginal parameters in the presence of informative cluster sizes.

The WCR method randomly samples one representative observation from each of I clusters with replacement. The resampled data set (size I) is analyzed by a generalized linear model, e.g. logistic model, since the observations within a resampled data set are independent. This process is repeated a large number of times, say Q, where each of the Q analyses provides a consistent estimate of the effect of interest. Conditional on the data, the estimates from the q^{th} resample, $\underset{\sim}{\widehat{\beta}}(R;q)$, $q = 1, \ldots, Q$, are independent and identically distributed. However, since the resampled data sets contain correlated and overlapping observations, the Q resampled data sets are dependent. The dependency does not affect the WCR estimator, but it is implicitly taken into account by the variance formula. The overall estimate of the effect of interest is found by computing the average of the Q resample-based estimates. Figure 5.1 provides a visual representation of the WCR sampling scheme.

Figure 5.1. WCR Sampling Scheme

Consider a hypothetical data set consisting of I clusters ...

$$\underset{\sim}{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \end{bmatrix} \quad \underset{\sim}{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \end{bmatrix} \quad \underset{\sim}{y}_3 = [y_{31}] \quad \underset{\sim}{y}_4 = \begin{bmatrix} y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{bmatrix} \quad \cdots \quad \underset{\sim}{y}_I = \begin{bmatrix} y_{I1} \\ y_{I2} \end{bmatrix}$$

By repeated (Q times), random sampling of one observation from each of the above I clusters, a series of Q, $q = 1, \ldots, Q$, resampled data sets $R(q)$ of size I and resample-based estimates $(\underset{\sim}{\hat{\beta}}\,(R;q))$ are obtained ...

$$R(1) = \left| \begin{matrix} y_{11}, \underset{\sim}{x}_{11} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{44}, \underset{\sim}{x}_{44} \\ \vdots \\ y_{I1}, \underset{\sim}{x}_{I1} \end{matrix} \right| \underset{\rightarrow}{\text{GLM}}\ \underset{\sim}{\hat{\beta}}\,(R;1) \quad \cdots \quad R(Q) = \left| \begin{matrix} y_{12}, \underset{\sim}{x}_{12} \\ y_{22}, \underset{\sim}{x}_{22} \\ y_{31}, \underset{\sim}{x}_{31} \\ y_{42}, \underset{\sim}{x}_{42} \\ \vdots \\ y_{I2}, \underset{\sim}{x}_{I2} \end{matrix} \right| \underset{\rightarrow}{\text{GLM}}\ \underset{\sim}{\hat{\beta}}\,(R;Q)$$

The WCR estimator $(\underset{\sim}{\bar{\beta}}\,)$ is obtained by averaging the resample-based estimates across the resamples ...

$$\underset{\sim}{\bar{\beta}} = \frac{\sum_{q=1}^{Q} \underset{\sim}{\hat{\beta}}(R;q)}{Q}.$$

A consistent estimator for the WCR variance, $\widehat{Var}(\underset{\sim}{\bar{\beta}}\,)$, can be written as

$$\underset{\approx}{\hat{\Sigma}} = \widehat{Var}(\underset{\sim}{\bar{\beta}}\,) \doteq \frac{\sum_{q=1}^{Q} \underset{\approx}{\hat{\Sigma}}(R;q)}{Q} - \left(\frac{Q-1}{Q}\right) \underset{\approx}{S}^2_{\underset{\sim}{\beta}},$$

where $\widehat{\underset{\sim}{\Sigma}}\ (\boldsymbol{R};\boldsymbol{q})$ is the estimated covariance matrix from the q^{th} analysis and

$$\underset{\sim}{S}^{2}_{\underset{\sim}{\beta}} = \frac{\sum_{q=1}^{Q}\underset{\sim}{\widehat{\beta}}(R;q)\underset{\sim}{\widehat{\beta}}(R;q)^{T} - Q\left(\frac{\sum_{q=1}^{Q}\underset{\sim}{\widehat{\beta}}(R;q)}{Q}\right)\left(\frac{\sum_{q=1}^{Q}\underset{\sim}{\widehat{\beta}}(R;q)}{Q}\right)^{T}}{Q-1}$$

is the covariance matrix among the $\underset{\sim}{\widehat{\beta}}\ (R;q)$'s.

5.4 Simulation Results

5.4.1 Introduction

Informative cluster size simulations are important because they demonstrate the finite sample behavior in situations where the WCR method is valid and the GEE method is not. Informative cluster sizes can be accomplished by forcing clusters with a lower than usual baseline risk to be larger in size: their cluster size is taken as a truncated, $binomial(0.75, 9)$, excluding the 0's and 9's. On the other hand, clusters with a higher than usual baseline risk are forced to be smaller: their cluster size is taken as a truncated, $binomial(0.25, 9)$, excluding the 0's and 9's. The simulations presented in this chapter are set up in the same way as those in Chapter IV with null and positive exposure effects. Table 5.1 demonstrates that GEE is not a valid or consistent method for data with informative cluster sizes. The simulated example has observation-specific covariates, $\rho = 0.2$, $n_i \leq 8$, and $I = 250$.

Table 5.1. Positive Exposure Effect with Informative Cluster Sizes, $\rho = 0.2$ (OS)

	WCR	GEE(r)	GEE(mb)	GEE(r)	GEE(mb)
		independence		exchangeable	
S	1000	1000	1000	1000	1000
Q	50,000				
I	250	250	250	250	250
n_i	≤ 8	≤ 8	≤ 8	≤ 8	≤ 8
$\overline{b}_0$	-1.1101	-1.5053	-1.5053	-1.3652	$-.13652$
$ese(\widehat{b}_0)$	0.0046	0.0039	0.0039	0.0041	0.0041
$\overline{b}_1$	0.4907	0.5017	0.5017	0.4835	0.4835
$ese(\widehat{b}_1)$	0.0057	0.0046	0.0046	0.0042	0.0042
$std(\widehat{b}_0)$	0.1402	0.1191	0.1060	0.1144	0.1114
$std(\widehat{b}_1)$	0.1735	0.1404	0.1405	0.1255	0.1274

$b_0 = -1.0986; b_1 = 0.4796.$

It can be seen by this simulated scenario that WCR and GEE are not estimating the same parameters. Also, GEE under independence and GEE under exchangeable variance structures are not estimating the same parameters. This is cause for great concern since the user-specified working correlation structure should not affect the value of the parameter estimates, but only affect the efficiency of the estimation. It is difficult to determine what parameter GEE is estimating and what the proper interpretation is under informative cluster sizes. The appropriateness of current methods in the presence of informative cluster sizes will be researched in more detail in the future. However, further comparison of WCR and GEE will not be done since GEE clearly is not a viable choice for analyzing informative cluster size data.

Simulations similar in design to those in Chapter IV were performed under a variety of scenarios with informative cluster sizes. Table 5.2 displays the specifications for the simulation results discussed in this chapter.

Table 5.2. Informative Cluster Size Simulation Specifications

Table	Exposure	Covariate	ρ	Q	I	n_i	b_0	b_1
5.3	Null	CS	0.2	10,000	250	≤ 8	−1.0986	0.0000
5.3	Null	CS	0.4	10,000	250	≤ 8	0.6931	0.0000
5.4	Positive	CS	0.2	10,000	50 − 250	≤ 8	−1.0986	0.4796
5.5	Positive	CS	0.4	10,000	50 − 250	≤ 8	0.6931	0.4796
5.6	Positive	OS	0.2	10,000	250	≤ 8	−1.0986	0.4796
5.7	Positive	OS	0.4	10,000	250	≤ 8	0.6931	0.4796

5.4.2 Cluster-specific Covariates

Informative cluster sizes can be accomplished by forcing clusters with a lower than usual baseline risk to be larger in size as described in Section 5.4.1. The simulations presented in this chapter are designed in the same way as those in Chapter IV with null and positive exposure effects. The true values for the intercept parameter for the WCR analyses are the same as if the data did not have informative cluster sizes. The simulation results presented in the rest of this chapter will only be for WCR because GEE is not appropriate for data with informative cluster sizes. The results of the informative cluster size simulations using a null exposure effect with cluster-specific covariates are displayed in Table 5.3.

Table 5.3. Null Exposure Effect with Informative Cluster Sizes (CS)

S	1000	1000
Q	10,000	10,000
I	250	250
n_i	≤ 8	≤ 8
ρ	0.2	0.4
$\overline{\hat{b}_0}$	-1.1099	0.6984
$ese(\hat{b}_0)$	0.0048	0.0047
$\overline{\hat{b}_1}$	0.0048	0.0050
$ese(\hat{b}_1)$	0.0068	0.0066
$std(\hat{b}_0)$	0.1485	0.1451
$std(\hat{b}_1)$	0.2102	0.2054
coverage	0.943	0.944
empirical alpha	0.057	0.056

$\rho = 0.2 : b_0 = -1.0986, b_1 = 0.0000;$
$\rho = 0.4 : b_0 = 0.6931, b_1 = 0.0000.$

The intercept parameter estimate for $\rho = 0.2$ is slightly biased, whereas the other parameter estimates are not. The average coverage and empirical alpha of b_1 for both $\rho = 0.2$ and 0.4 are good.

Positive exposure effect scenarios using cluster-specific exposures with informative cluster sizes are accomplished in the same way as done in Section 4.3 of Chapter IV. The true values for the parameters for the WCR analyses remain the same, $b_0 = -1.0986$ and $b_1 = 0.4796$ when $\rho = 0.2$, and $b_0 = 0.6931$ and $b_1 = 0.4796$ when $\rho = 0.4$. The results of the positive exposure effect simulations using cluster-specific covariates with informative cluster size are displayed in Tables 5.4-5.7.

Table 5.4. Positive Exposure Effect with Informative Cluster Sizes, $\rho = 0.2$ (CS)

S	1000	1000	1000
Q	10,000	10,000	10,000
I	50	100	250
n_i	≤ 8	≤ 8	≤ 8
$\overline{\hat{b}}_0$	-1.1497	-1.1344	-1.1126
$ese(\hat{b}_0)$	0.0118	0.0076	0.0048
$\overline{\hat{b}}_1$	0.4811	0.4996	0.4922
$ese(\hat{b}_1)$	0.0154	0.0103	0.0066
$\overline{std(\hat{b}_0)}$	0.3387	0.2349	0.1488
$\overline{std(\hat{b}_1)}$	0.4554	0.3178	0.2000
coverage	0.943	0.944	0.938
power	0.211	0.356	0.701
# zero cells out of $S \cdot Q$	13,059	17	0

$b_0 = -1.0986; b_1 = 0.4796.$

Table 5.5. Positive Exposure Effect with Informative Cluster Sizes, $\rho = 0.4$ (CS)

S	1000	1000	1000
Q	10,000	10,000	10,000
I	50	100	250
n_i	≤ 8	≤ 8	≤ 8
$\overline{\hat{b}}_0$	0.7305	0.7096	0.6984
$ese(\hat{b}_0)$	0.0109	0.0076	0.0046
$\overline{\hat{b}}_1$	0.5225	0.4935	0.4890
$ese(\hat{b}_1)$	0.0155	0.0110	0.0066
$\overline{std(\hat{b}_0)}$	0.3288	0.2306	0.1449
$\overline{std(\hat{b}_1)}$	0.4834	0.3363	0.2119
coverage	0.953	0.947	0.957
power	0.191	0.331	0.636
# zero cells out of $S \cdot Q$	18,637	38	0

$b_0 = 0.6931; b_1 = 0.4796;$
† indicates more than 2 standard deviations from nominal value, 0.95.

All of the intercept parameter estimates and the exposure parameter estimate for $I = 50$ are slightly biased for $\rho = 0.2$. For $\rho = 0.4$, the intercept parameter estimates for $I = 50$ and 100 are slightly biased, as well as the exposure parameter estimate for $I = 50$. The WCR coverages are good and the power increases nicely with the sample size.

5.4.3 Observation-specific Covariates

Observation-specific exposures were also simulated. The set up is exactly the same as the informative cluster size simulations with cluster-specific exposure. The observation-specific cases were simulated for low ($\rho = 0.2$) and moderate ($\rho = 0.4$) within-cluster correlation cases with $I = 250$. More resamples were necessary to obtain stable parameter and variance estimates. Simulation results are given in Tables 5.6-5.7.

Table 5.6. Positive Exposure Effect with Informative Cluster Sizes, $\rho = 0.2$ (OS)

S	1000
Q	50,000
I	250
n_i	≤ 8
$\overline{b}_0$	-1.1101
$ese(\widehat{b}_0)$	0.0046
$\overline{b}_1$	0.4907
$ese(\widehat{b}_1)$	0.0057
$\overline{std(\widehat{b}_0)}$	0.1402
$\overline{std(\widehat{b}_1)}$	0.1735
coverage	0.936
power	0.814

$b_0 = -1.0986; b_1 = 0.4796.$

Table 5.7. Positive Exposure Effect with Informative Cluster Sizes, $\rho = 0.4$ (OS)

S	1000
Q	50,000
I	250
n_i	≤ 8
$\overline{b}_0$	0.7079
$ese(\widehat{b}_0)$	0.0041
$\overline{b}_1$	0.4844
$ese(\widehat{b}_1)$	0.0050
$\overline{std(\widehat{b}_0)}$	0.1274
$\overline{std(\widehat{b}_1)}$	0.1520
coverage	0.937
power	0.881

$b_0 = 0.6931; b_1 = 0.4796.$

The intercept parameter estimates are again slightly biased for WCR under informative cluster sizes, although the exposure parameter estimates are not. WCR slightly under covered for $\rho = 0.2$, which indicates that $Q = 50,000$ might not be large enough. However, the two simulations had good power as shown in Tables 5.6 and 5.7.

5.5 Q-Q Plots

Q-Q plots also were done for each WCR simulation of Chapter V. The Q-Q plots show the 2000 standardized parameter estimates from a WCR simulation plotted against the quantiles of the standard normal distribution. If the data points fall on a 45° line, then the data suggest that the distribution of the parameter is standard normal. The Q-Q plots proved visual evidence of the asymptotic normality of a standardized WCR estimator. A Q-Q plot was done for both the intercept and the exposure parameter for all of the simulations in Chapter V. The Q-Q plot is titled according to which table it corresponds with additional information in a sub-title when more than one simulation is presented in a table in Chapter V. The Q-Q plots corresponding to the simulations of Chapter V can be found in Appendix 1. As can be seen by the Q-Q plots in Appendix 1, the simulations strongly suggest that the parameter estimates are asymptotically normal.

5.6 Example

Even though GEE is not a valid analysis method with informative cluster size data, many data analysts will employ GEE in spite of this fact. Also, many times it will be unclear whether informative cluster sizes do exist. For these reasons, we have analyzed the Paul (1982) litter-effects data set using the WCR and GEE methods.

Table 5.8. Data from Paul (1982)

Control	1 1 4 0 0 0 0 0 1 0 2 0 5 2 1 2 0 0 1 0 0 0 0 3 2 4 0
	12 7 6 6 7 8 10 7 8 6 11 7 8 9 2 7 9 7 11 10 4 8 10 12 8 7 8
Low	0 1 1 0 2 0 1 0 1 0 0 3 0 0 1 5 0 0 3
	5 11 7 9 12 8 6 7 6 4 6 9 6 7 5 9 1 6 9
Medium	2 3 2 1 2 3 0 4 0 0 4 0 0 6 6 5 4 1 0 3 6
	4 4 9 8 9 7 8 9 6 4 6 7 3 13 6 8 11 7 6 10 6
High	1 0 1 0 1 0 1 1 2 0 4 1 1 4 2 3 1
	9 10 7 5 4 6 3 8 5 4 4 5 3 8 6 8 6

*Within each treatment group the top number is the number of pups affected by treatment and the number below it is the total number of pups in the litter.

Both analyses of the data were done with a dichotomous treatment variable: control vs. treated with the treated groups combined (model is similar to (3.51)). A logistic model was fit to each resampled data set. A series of $Q = 2000$ resamples were performed in the WCR analysis, excluding the rare occurrence of an infinite estimate. The GEE analysis used a logit link with binomial variance and an exchangeable correlation structure. The choice of an exchangeable correlation structure is consistent with the usual assumption of equal correlation between all pups within a litter.

The analysis of the Paul data set also was done using ordinary logistic regression (OLR). This method is used to illustrate show that WCR and GEE do inflate the variances to account for the within-cluster correlation. OLR assumes that all of the observations within a litter are independent so it proceeds as if each observation contributes distinct and new information. Therefore, the variance estimates with positive within-cluster correlation from OLR tend to be underestimated, which results in inflated Type I error rates, i.e. an invalid test. The analysis results are as follows:

Table 5.9. PAUL (1982) ANALYSIS RESULTS

	WCR	GEE(r)	GEE(mb)	OLR
Q	2000			
I	84	84	84	84
n_i	$1-13$	$1-13$	$1-13$	$1-13$
$\widehat{\beta}_0$	-1.85	-1.86	-1.86	-1.86
$\widehat{\beta}_1$	0.679	0.714	0.714	0.714
$var(\widehat{\beta}_0)$	0.112	0.090	0.096	0.040
$var(\widehat{\beta}_1)$	0.156	0.120	0.130	0.054
Z_1	1.72	2.05	1.98	13.21

It is interesting to note that all of the approaches produce nearly the same parameter estimates. However, it is clear that OLR is not accounting for the within-cluster correlation since the variance estimates are much smaller than those from WCR.

One explanation for the slight difference in treatment parameter estimates between WCR and GEE is the possibility that cluster sizes are informative. As evidence of this note that the observation-based mean response is 0.171 for control litters of size ≤ 7 vs. 0.12 for litters of size > 7. Since informative cluster sizes may be present, GEE may not be a valid method. The conclusion drawn regarding the effect of the treatment in this data set does vary depending upon which analysis method is used. The WCR analysis does not provide evidence that there is a significant treatment effect, whereas, GEE finds a significant treatment effect at $\alpha = 0.05$.

5.7 Summary

An informative cluster size is defined as a relationship between the cluster size and the baseline risk. In a litter-effects study, if the larger litters tend to have a lower baseline risk than the smaller litters, then cluster size may be informative in this sense. This systematic difference can alter the value and interpretation of parameters depending upon

the analysis method used. Traditional marginal methods such as GEE increase the weight of the clusters with their size, and are not valid for data with informative cluster sizes.

The informative cluster size simulations verified that WCR has acceptable test size, confidence interval coverage and power. WCR performs well in estimating the cluster-based, marginal parameter for low and moderate levels of within-cluster correlation. It should have been obvious from the simulated scenario of Table 5.2 that GEE may not always provide consistent estimates when there are informative cluster sizes. Therefore, GEE could produce misleading results and WCR is the preferred analysis method when cluster sizes may be informative.

CHAPTER VI

CONCLUSIONS AND FUTURE WORK

Within Cluster Resampling has been shown to be a valid analysis method for clustered binary data. In Chapter III, we verified the asymptotic normality of the WCR estimator and established a consistent estimator for the WCR variance. The simulations in Chapter IV showed that WCR behaves well with finite samples. Also, with smaller sample sizes, the use of only the finite portion of the sampling distribution proved to have little adverse effect. The stability of the WCR parameter and variance estimates depends upon a large number of resamples. The simulations have shown that it is necessary for the number of resamples to be particularly large in the presence of low within-cluster correlation, large cluster sizes, or large sample sizes. In practice, one might first evaluate whether clustered methods are necessary at all. The number of resamples is adequate once stability of the parameter and variance estimates has been achieved. In the future, I will attempt to establish a stopping rule for the number of resamples.

WCR has certain advantages over GEE. It does not require a user-specified working correlation matrix. Also, WCR may be more appropriate than GEE for certain analyses. One nice feature of WCR is that it is a valid analysis method and estimates the same parameter regardless of whether there are informative or non-informative cluster sizes. Moreover, GEE requires that cluster sizes be non-informative while WCR has no such restriction.

Informative cluster sizes were discussed in Chapter V. They are a feature which differentiates the observation-based, marginal parameter from the cluster-based, marginal

parameter. The simulations in Chapter V demonstrate that WCR properly estimates parameters and has good coverage and power. Also, we have shown a scenario where GEE performs poorly in the informative cluster size situation reflecting the fact that GEE is not valid in such a context. Therefore, WCR is superior to GEE for analyzing informative cluster size data.

While working on Within Cluster Resampling, I have come across many interesting ideas for future work. Genetic family-based studies may an area of application for WCR. The repeated sampling of one member from each family is appealing since a genetic data set consists of a wide range of family relations. Therefore, GEE models are difficult to fit because the same correlation structure does not exist from family to family and relation to relation within family.

The random-effects equivalent of WCR, which we will term 'within cluster paired resampling' will also be researched. The idea is to repeatedly sample pairs of observations instead of just one observation from each cluster. In this way, a subject-specific model can be fit to each resampled data set. The sampling aspects, large sample theory, variance, and finite sample behavior will need to be explored.

Informative cluster sizes may also be an interesting area to investigate further. The incidence of this phenomenon in litter-effects studies as well as in other areas will be researched. The ramifications of informative cluster sizes on the assumptions and estimation procedures of analysis methods may be explored, as well as whether there are current methods that do not implicitly assume non-informative cluster sizes.

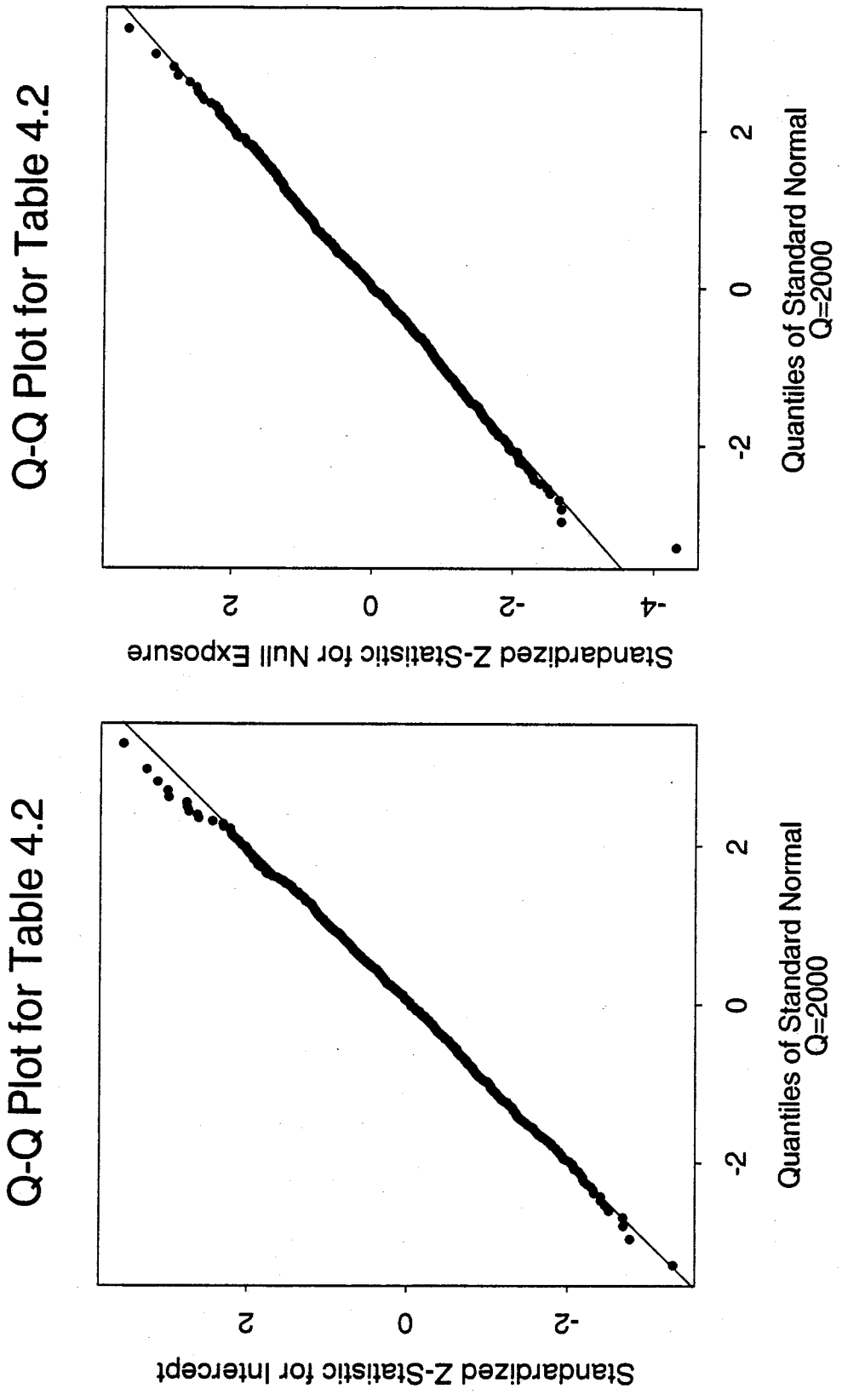
There are many issues relating to the source of the dependency among the outcomes of a cluster and heterogeneous response to exposure. Dependency can arise as a result of the presence of heterogeneity in the baseline. It can be a feature of the study design, such as auto-correlation in longitudinal studies; the effect of a contagion which infects one observation after another; or result from heterogeneous susceptibility, which means that every cluster or unit responds differently to an exposure. I will examine how

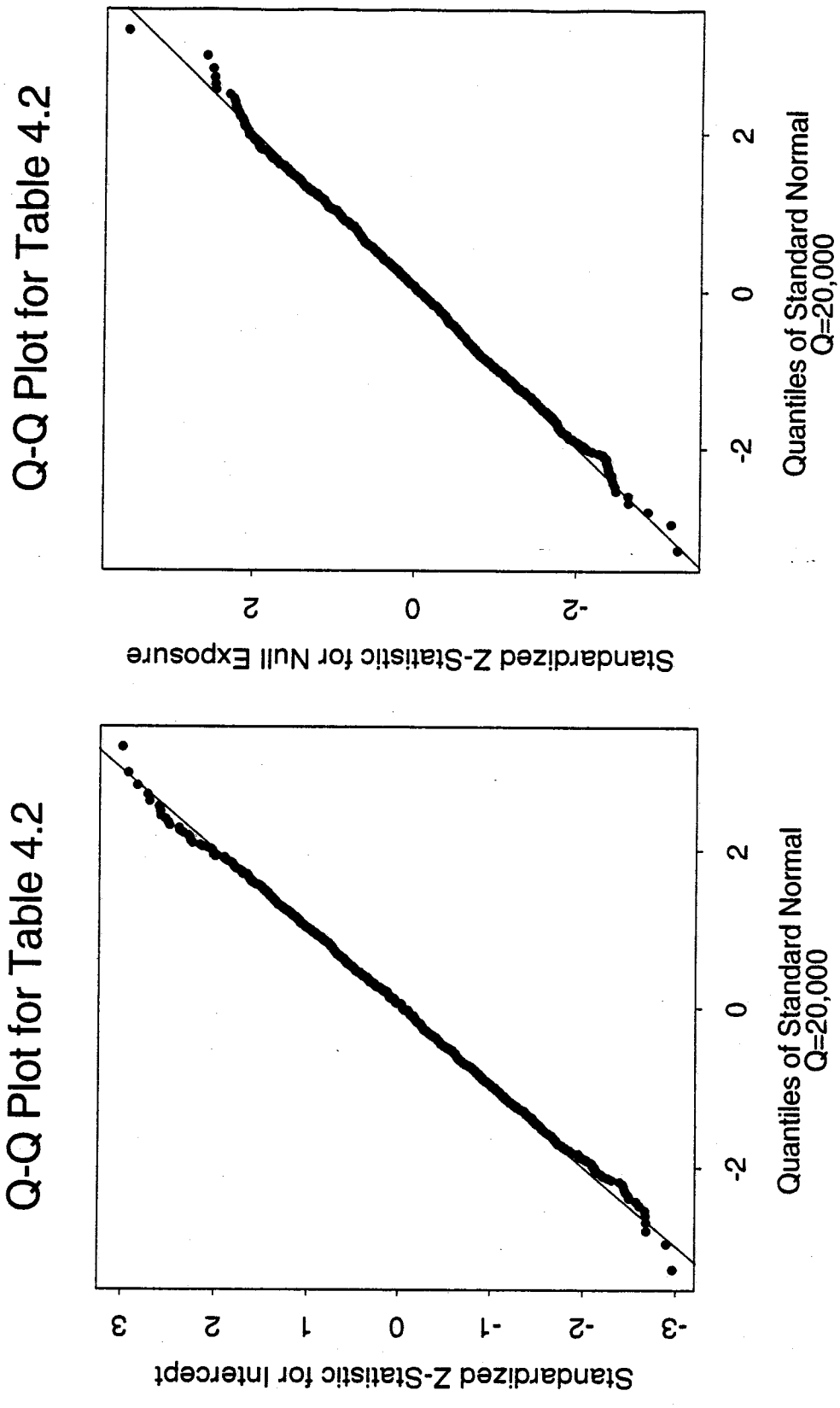
these sources of dependency interact and affect an analysis, as well as heterogeneous response to exposure.

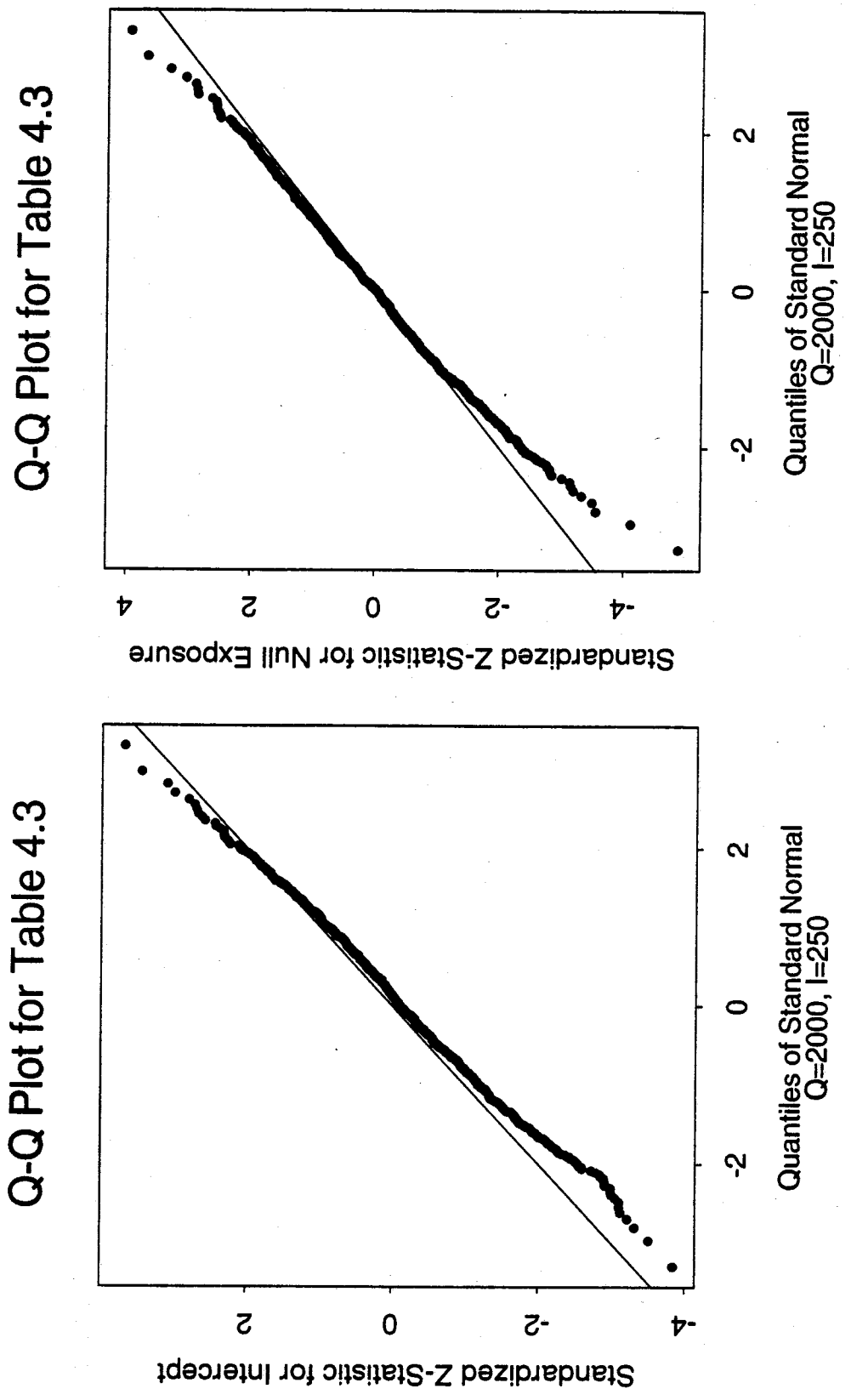
APPENDIX 1

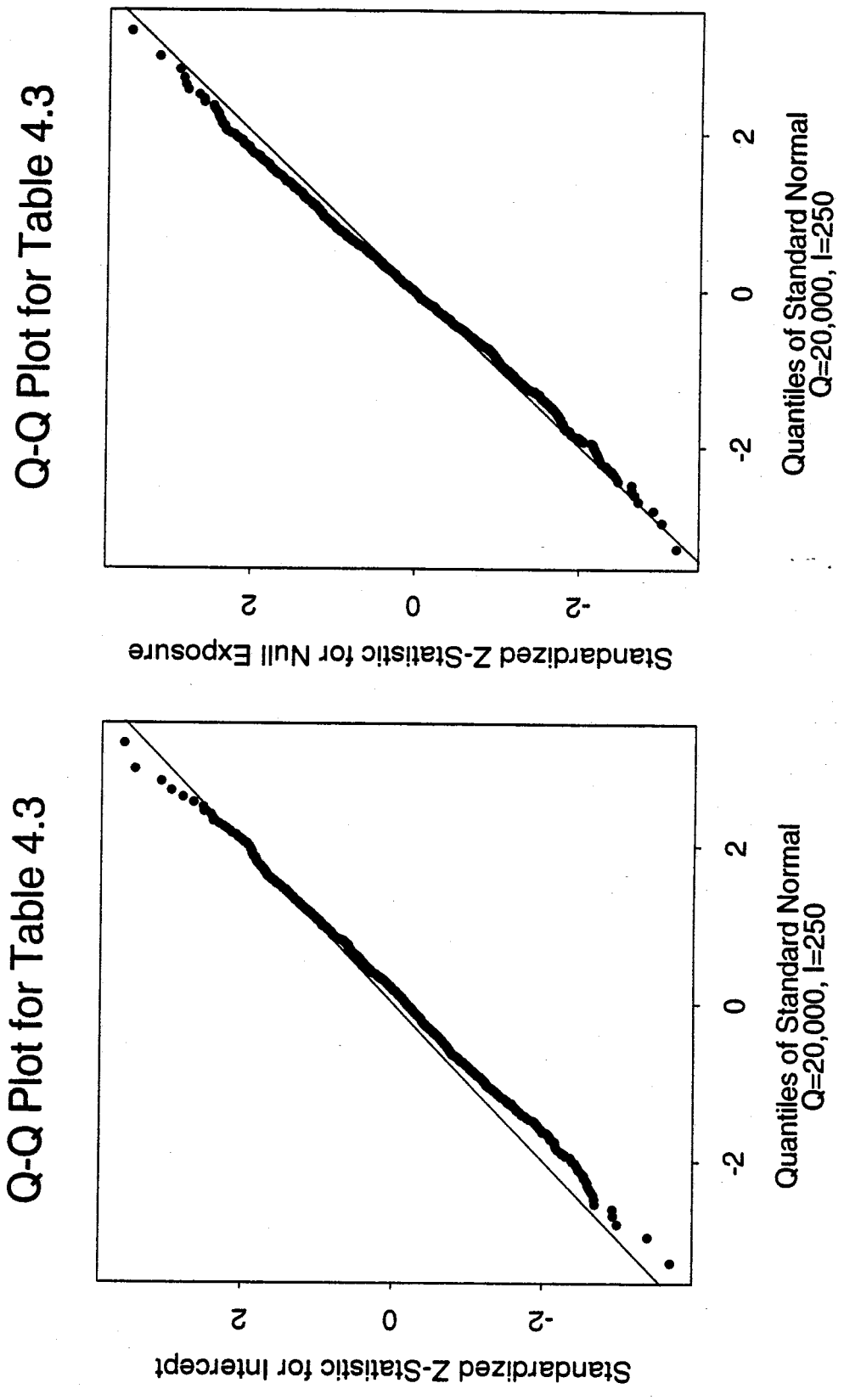
Q-Q PLOTS FOR SIMULATIONS IN CHAPTERS IV AND V

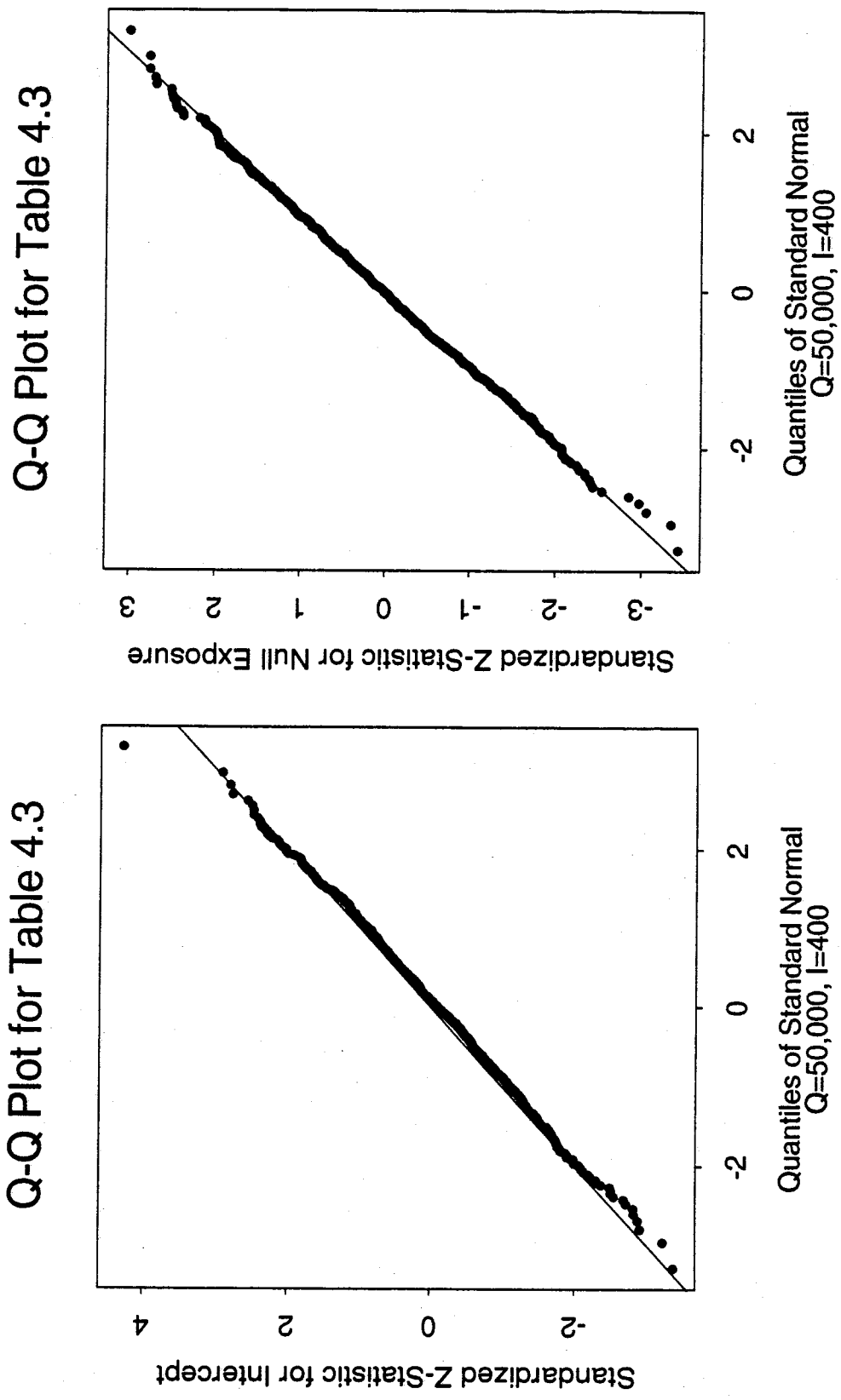
The Q-Q Plots that follow are the standardized parameter estimates from each of the 1000 simulated data sets plotted against the quantiles of the standard normal distribution. Two Q-Q Plots, one for the standardized intercept parameter and one for the standardized exposure parameter, are on a single page. They are labeled according to which table in Chapter IV or Chapter V they correspond with additional information to determine the specific simulation when more than one simulation result is presented in a single table.

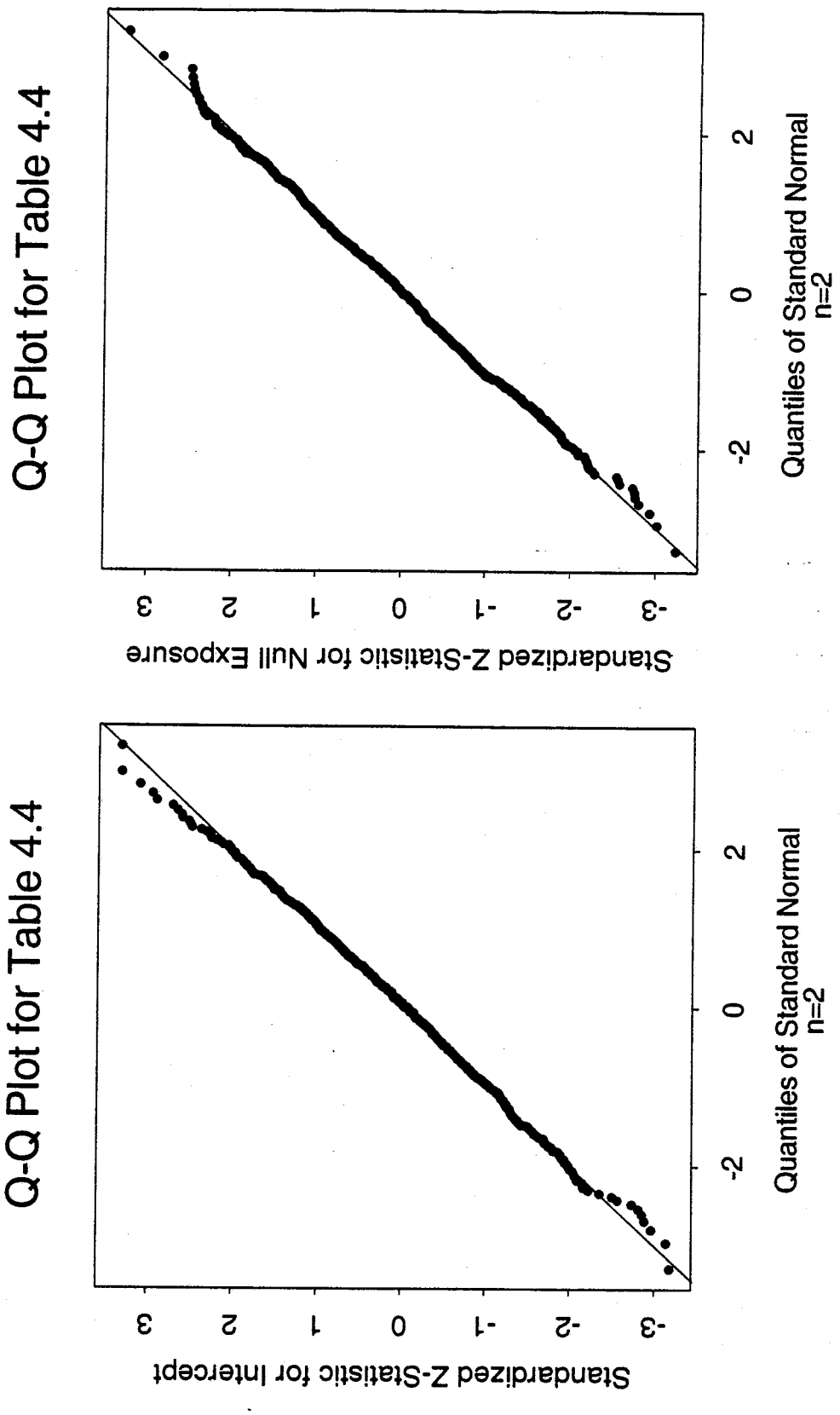

Q-Q Plot for Table 4.2
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q=2000
-2
0
2
Q-Q Plot for Table 4.2
Standardized Z-Statistic for Null Exposure
Quantiles of Standard Normal
Q=2000
-4
-2
0
2

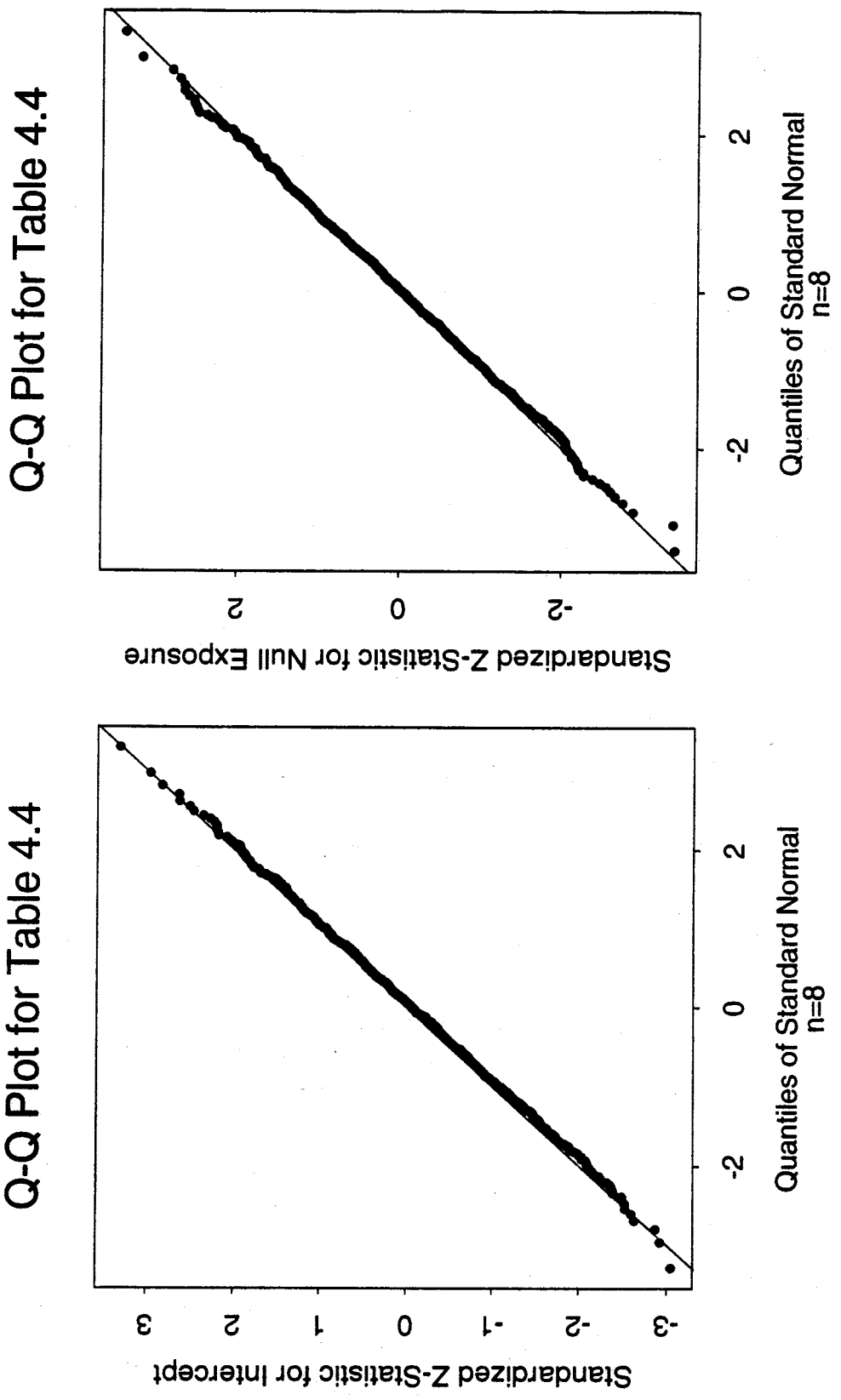

Q-Q Plot for Table 4.2
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=20,000
Q-Q Plot for Table 4.2
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=20,000

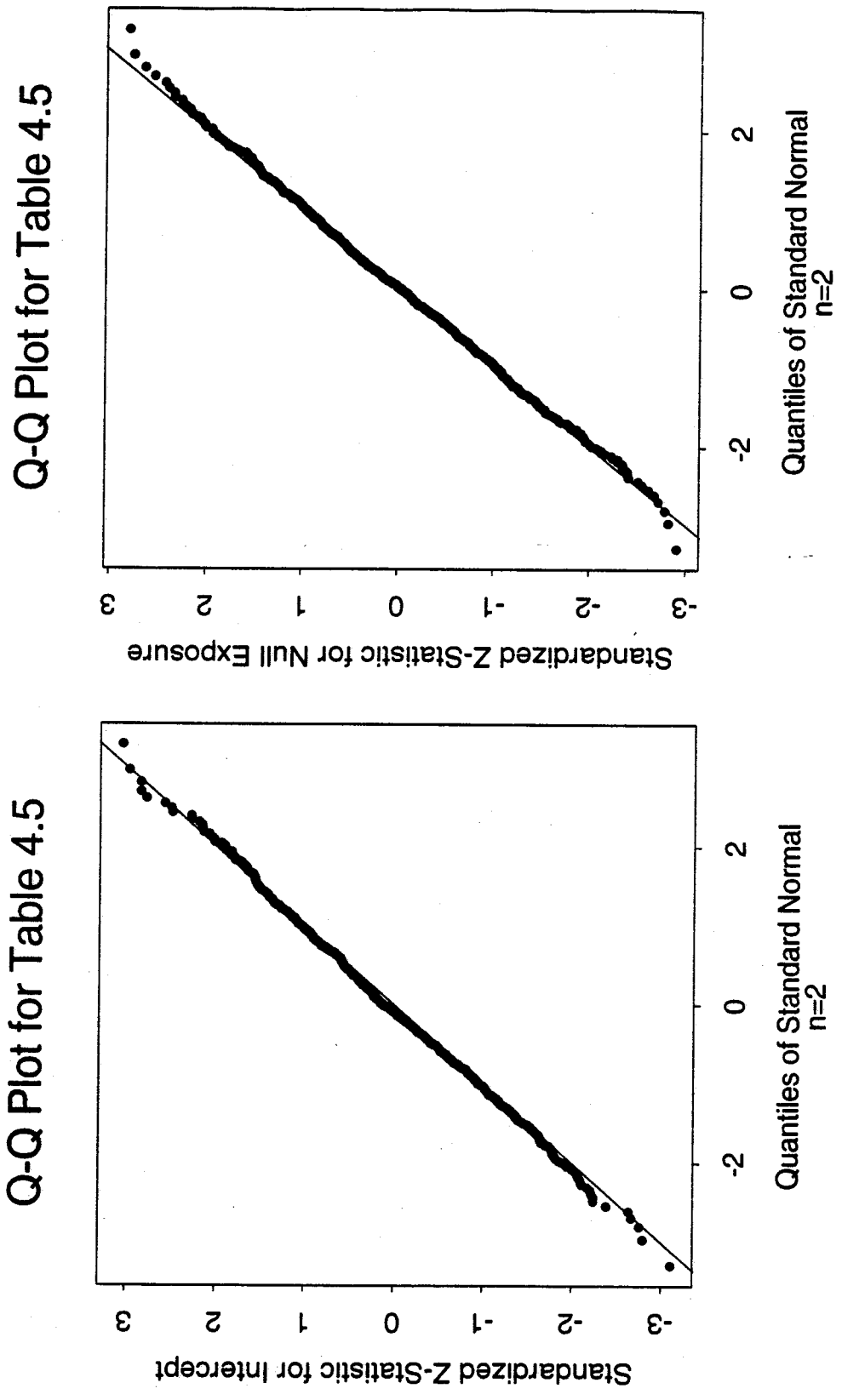

Q-Q Plot for Table 4.3
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=2000, I=250
Q-Q Plot for Table 4.3
Standardized Z-Statistic for Null Exposure
-4
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
Q=2000, I=250

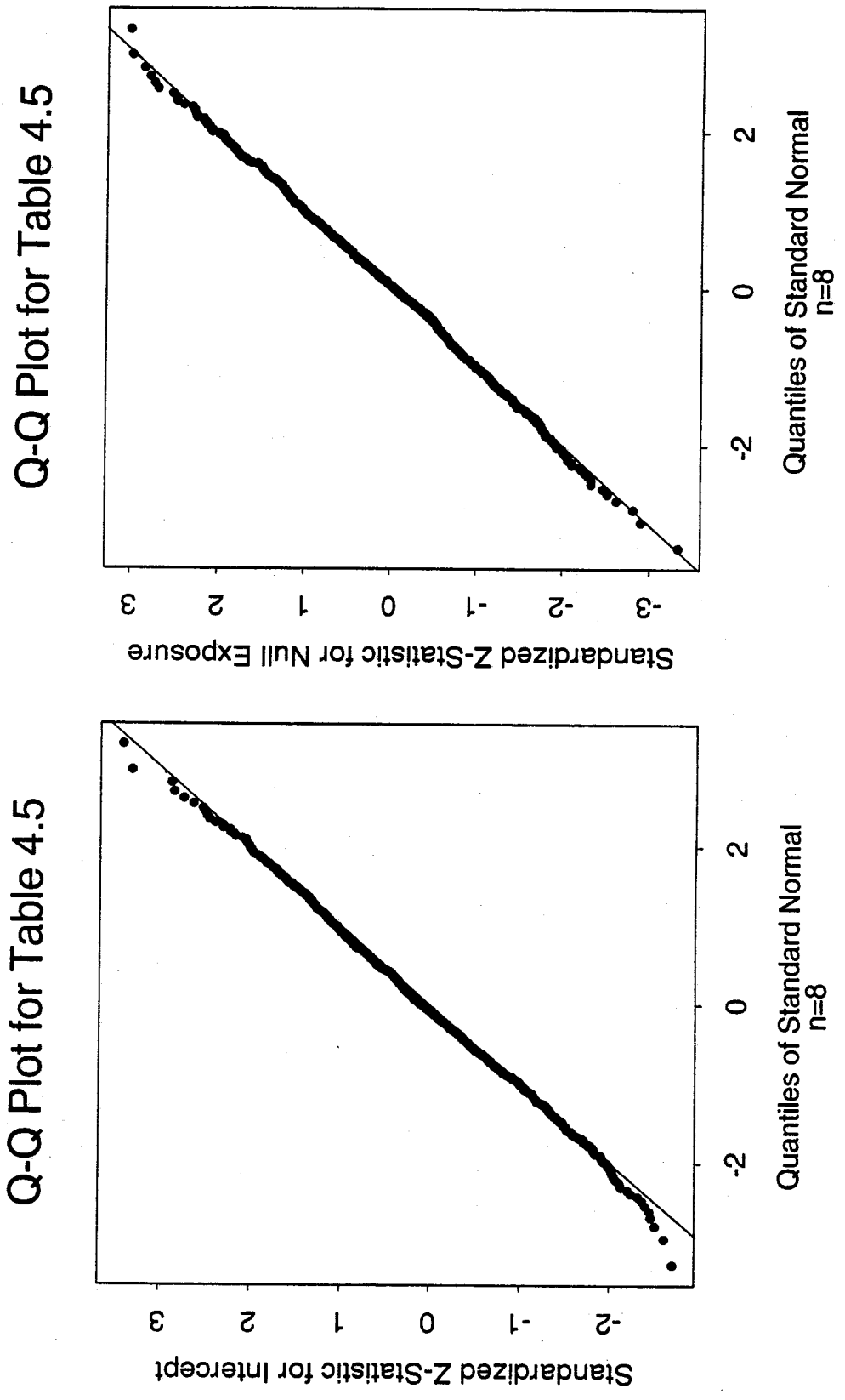

Q-Q Plot for Table 4.3
Standardized Z-Statistic for Intercept
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=20,000, I=250
Q-Q Plot for Table 4.3
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=20,000, I=250

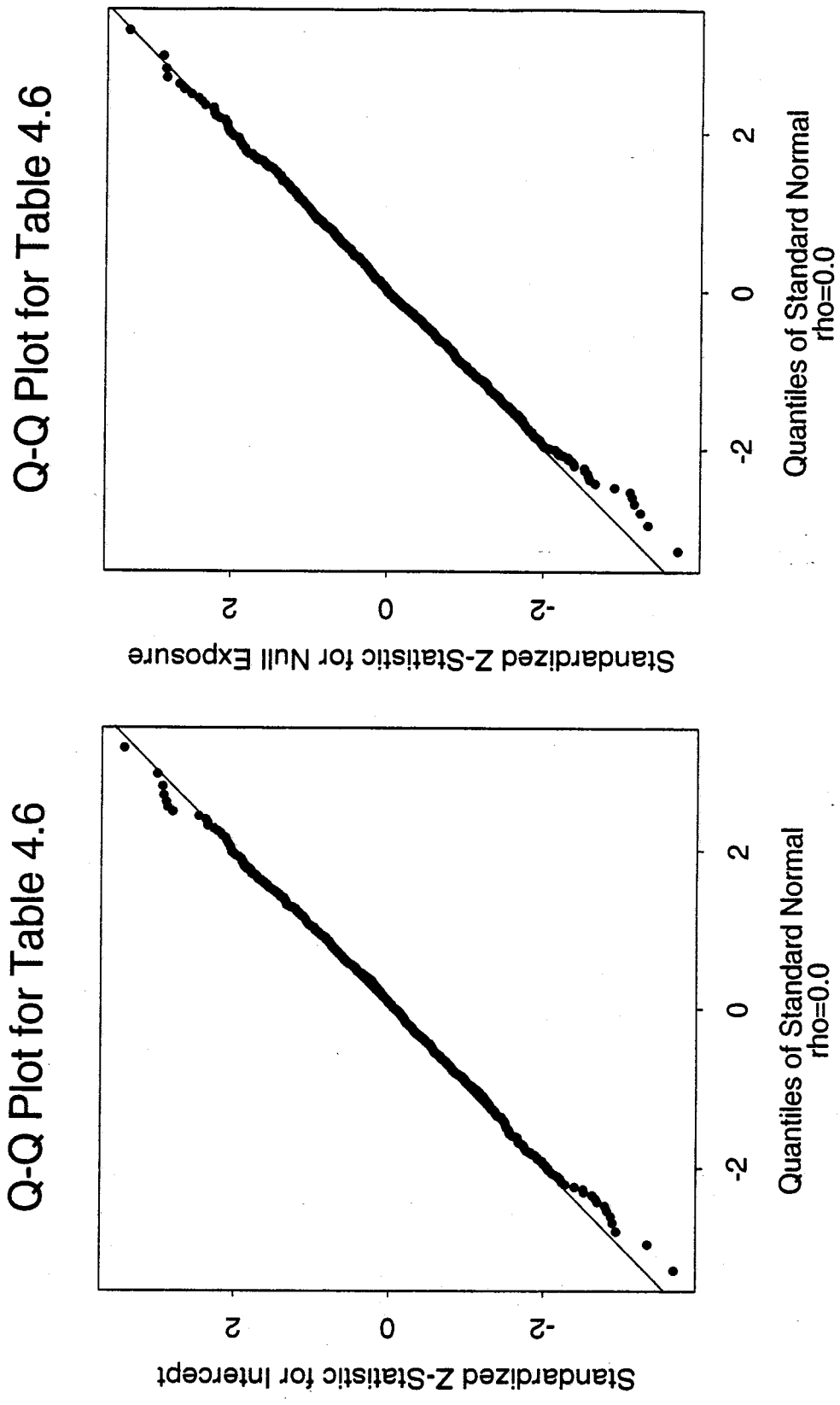

Q-Q Plot for Table 4.3
Standardized Z-Statistic for Intercept
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
Q=50,000, I=400
Q-Q Plot for Table 4.3
Standardized Z-Statistic for Null Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=50,000, I=400

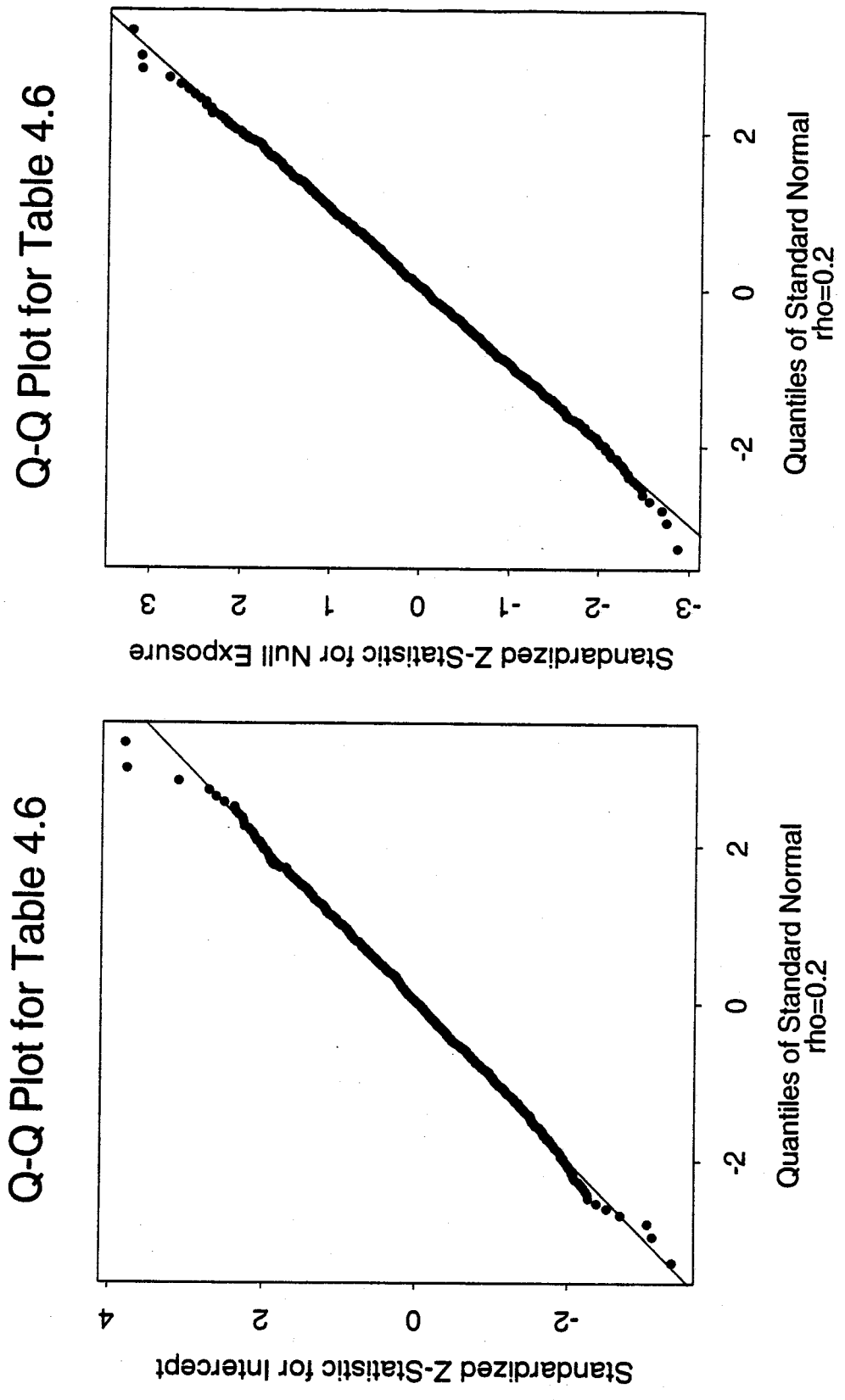

Q-Q Plot for Table 4.4
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
n=2
Q-Q Plot for Table 4.4
Standardized Z-Statistic for Null Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
n=2

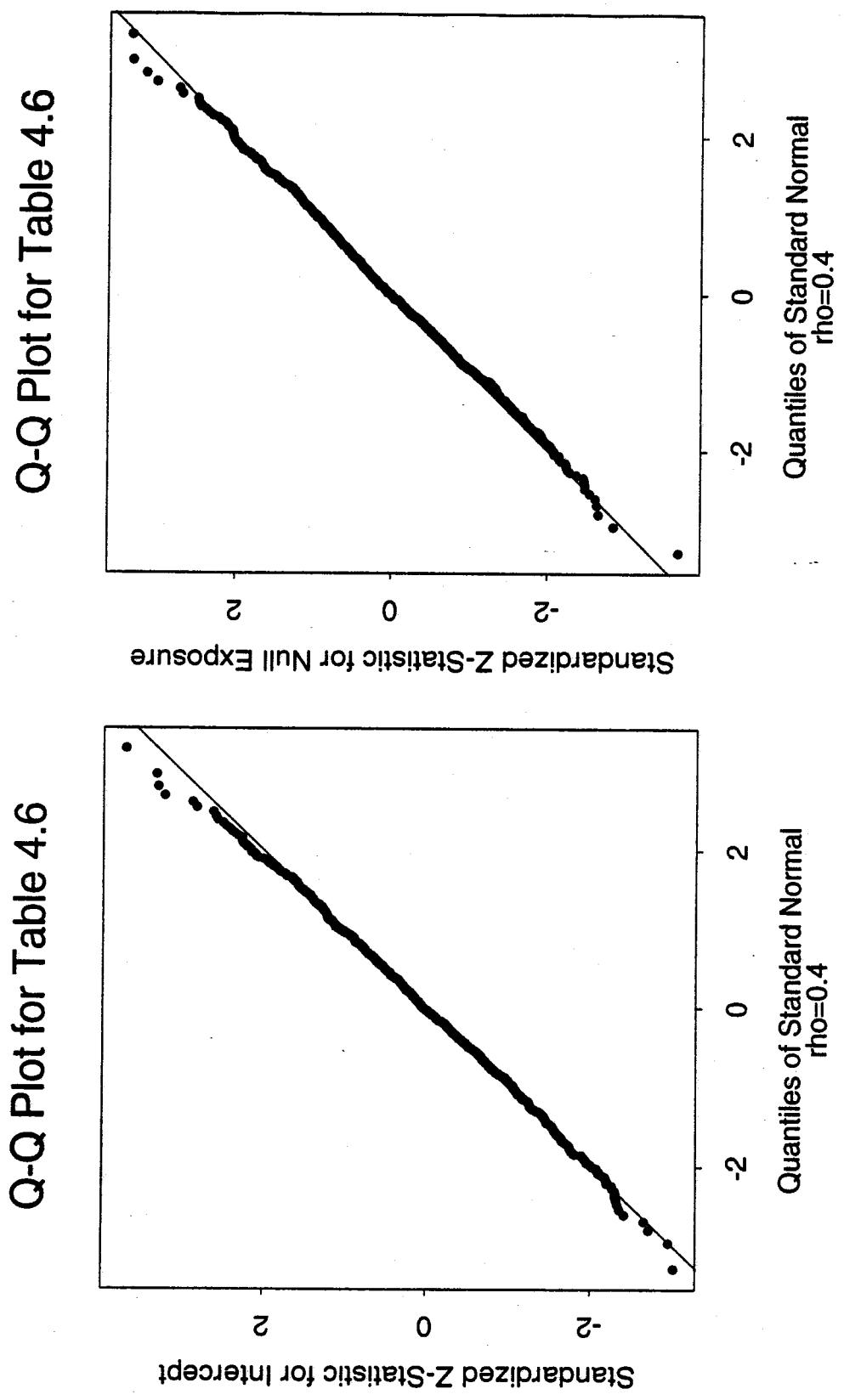

Q-Q Plot for Table 4.4
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
n=8
Q-Q Plot for Table 4.4
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
n=8

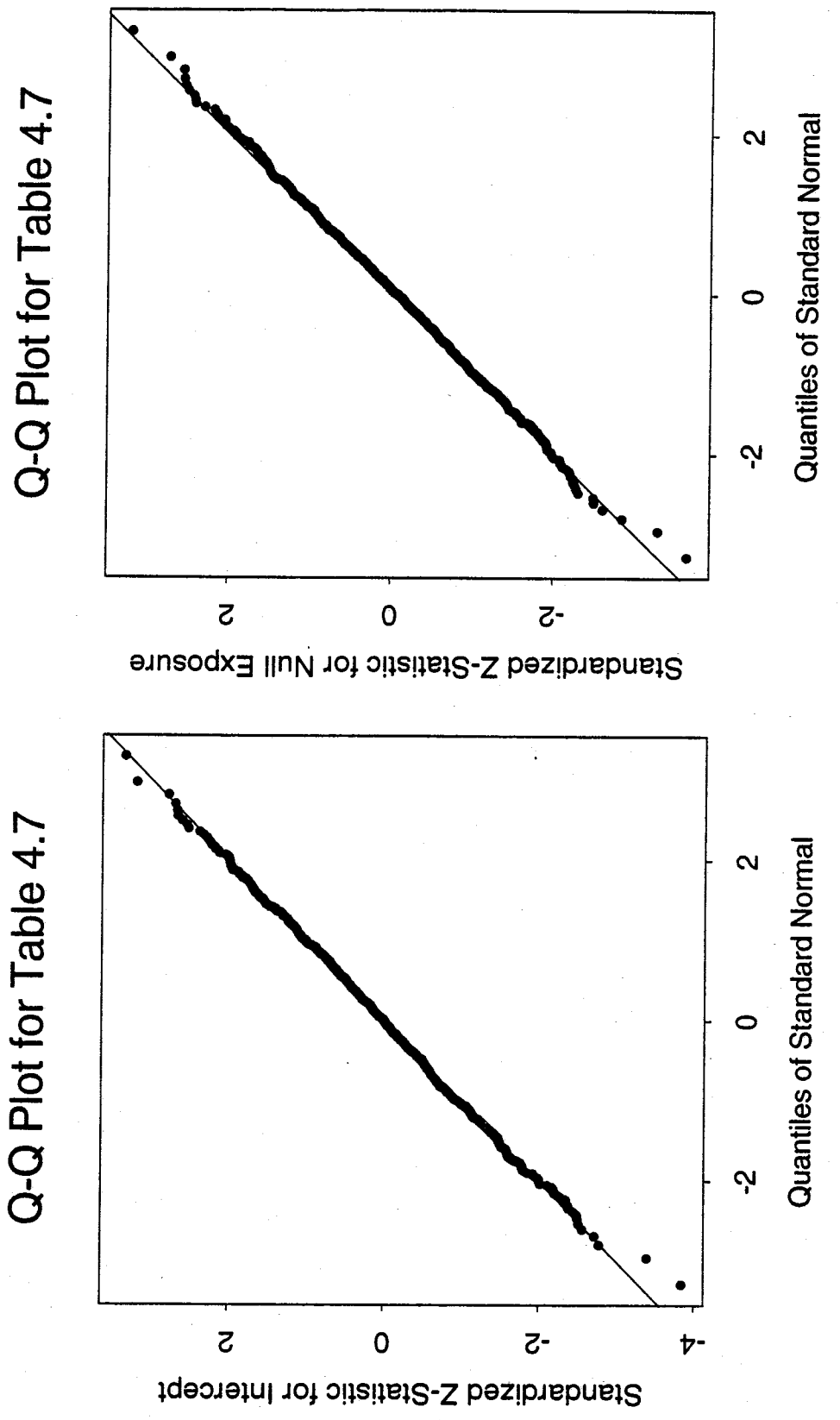

Q-Q Plot for Table 4.5
Standardized Z-Statistic for Null Exposure
Quantiles of Standard Normal
n=2
-3
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.5
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
n=2

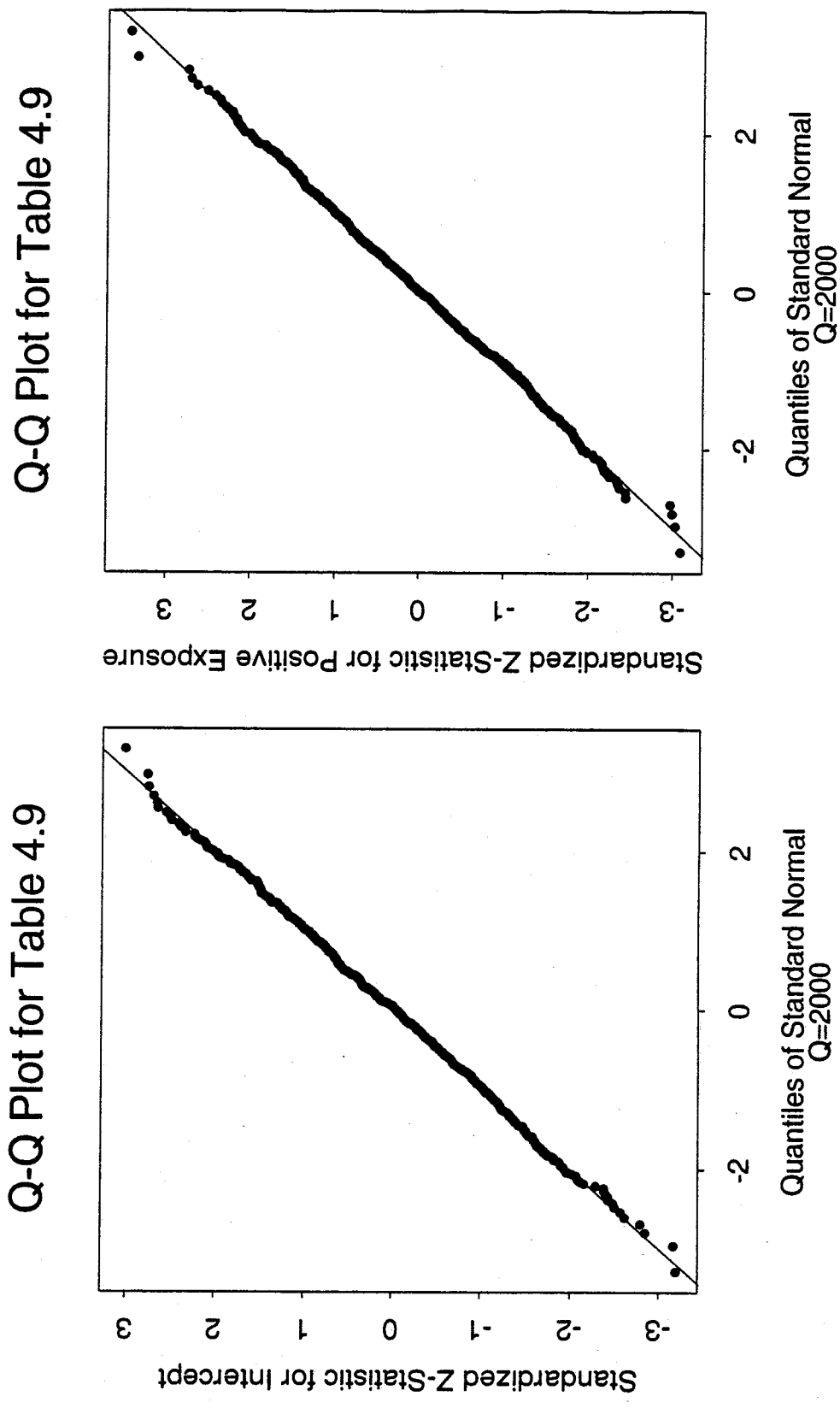

Q-Q Plot for Table 4.5
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
n=8
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.5
Standardized Z-Statistic for Null Exposure
Quantiles of Standard Normal
n=8
-3
-2
-1
0
1
2
3
-2
0
2

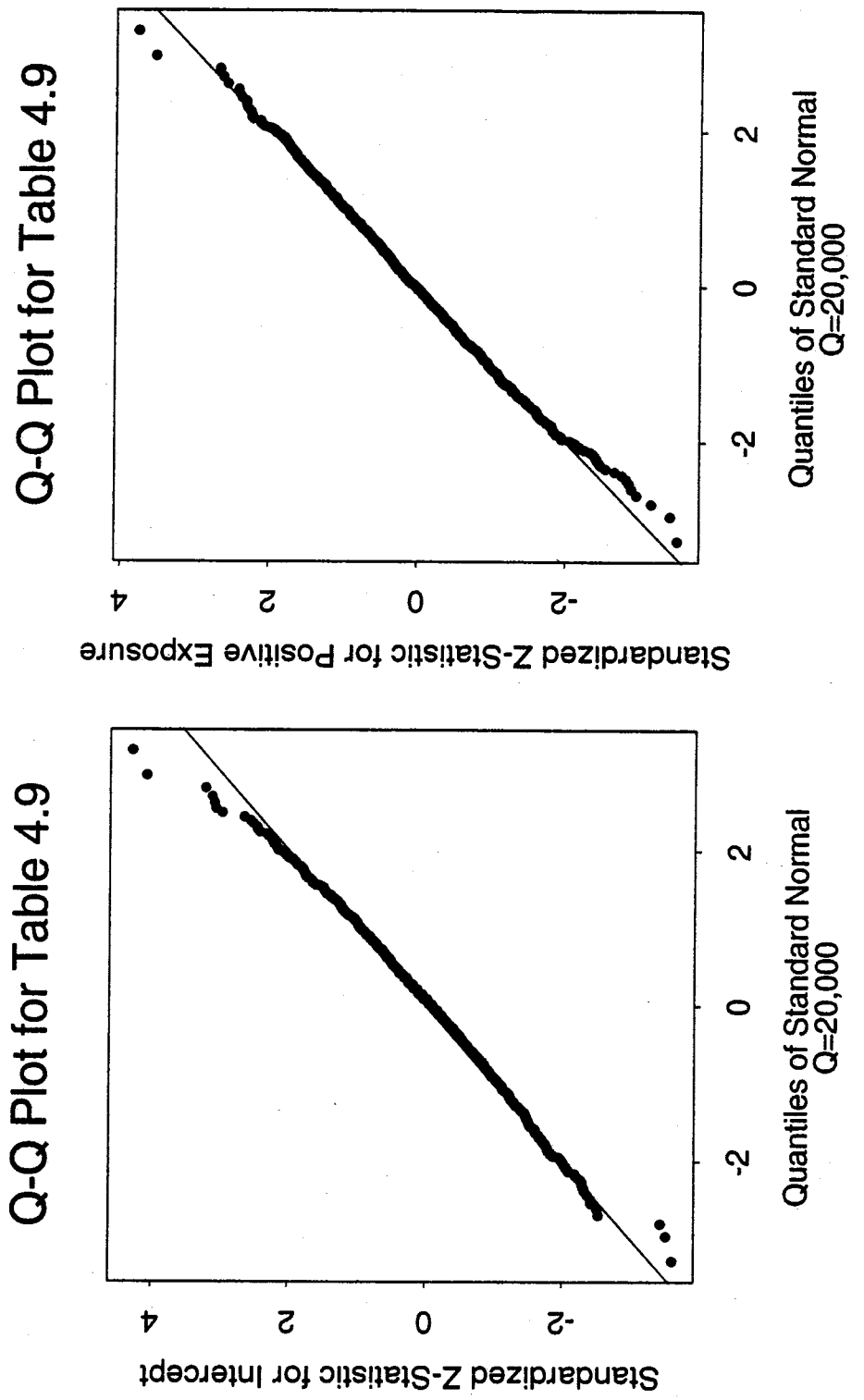

Q-Q Plot for Table 4.6
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
rho=0.0
Q-Q Plot for Table 4.6
Standardized Z-Statistic for Null Exposure
Quantiles of Standard Normal
rho=0.0

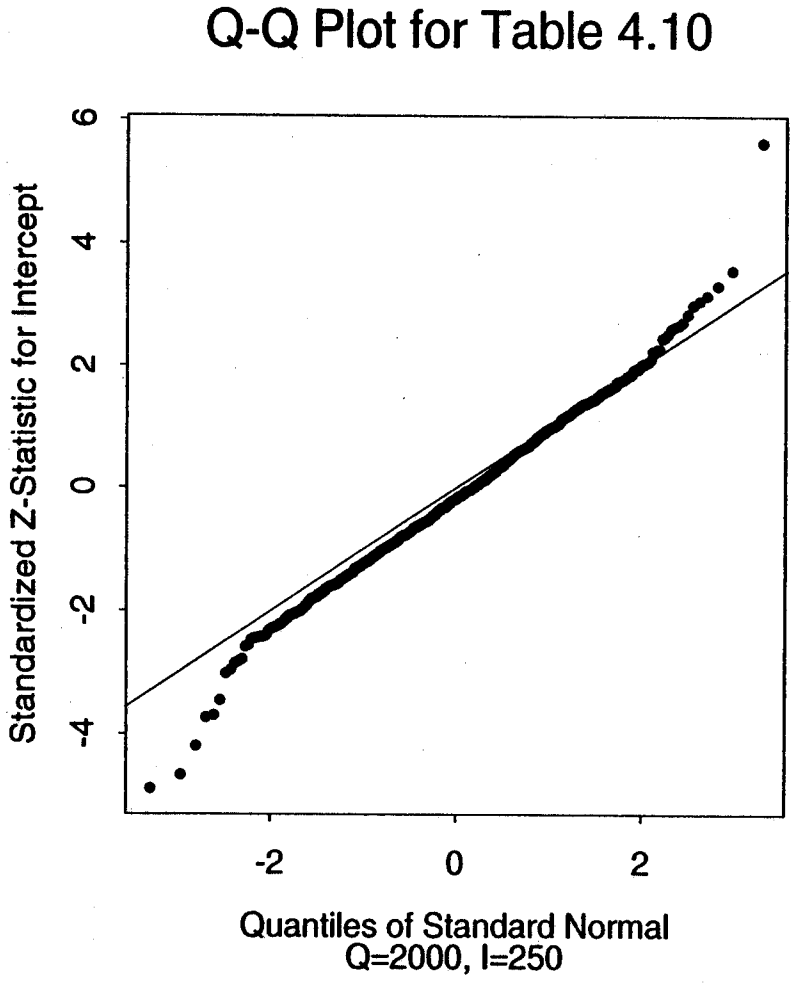

Q-Q Plot for Table 4.6
Standardized Z-Statistic for Intercept
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
rho=0.2
Q-Q Plot for Table 4.6
Standardized Z-Statistic for Null Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
rho=0.2

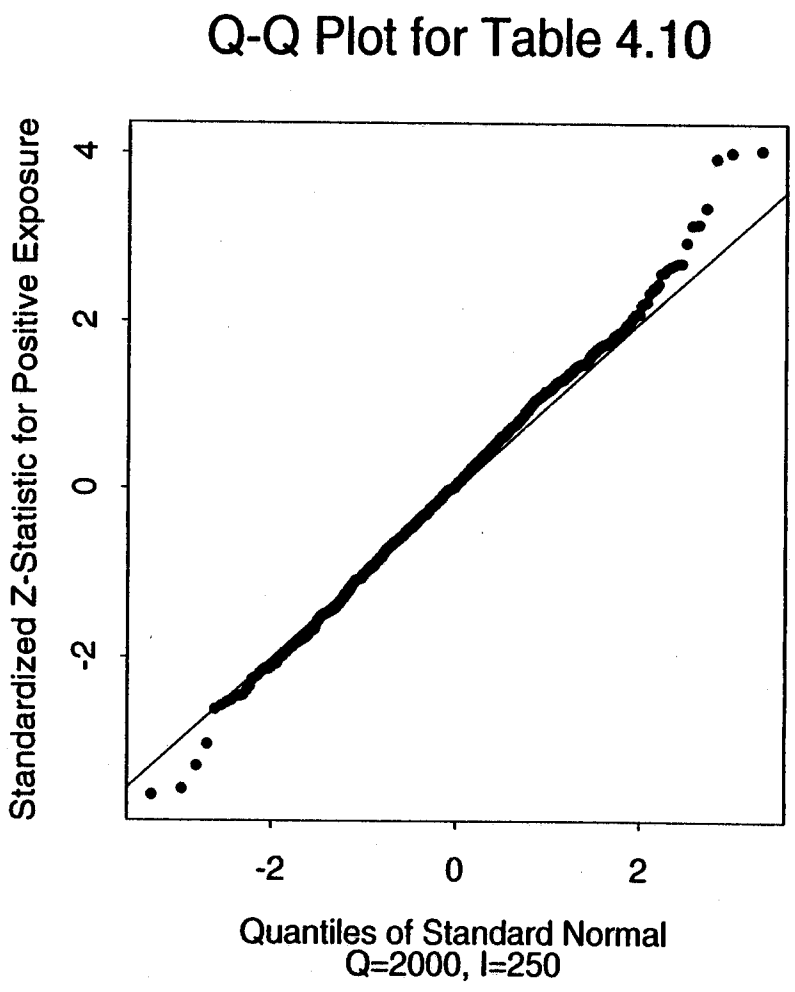

Q-Q Plot for Table 4.6
Standardized Z-Statistic for Intercept
-2
0
2
-2
0
2
Quantiles of Standard Normal
rho=0.4
Q-Q Plot for Table 4.6
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
rho=0.4

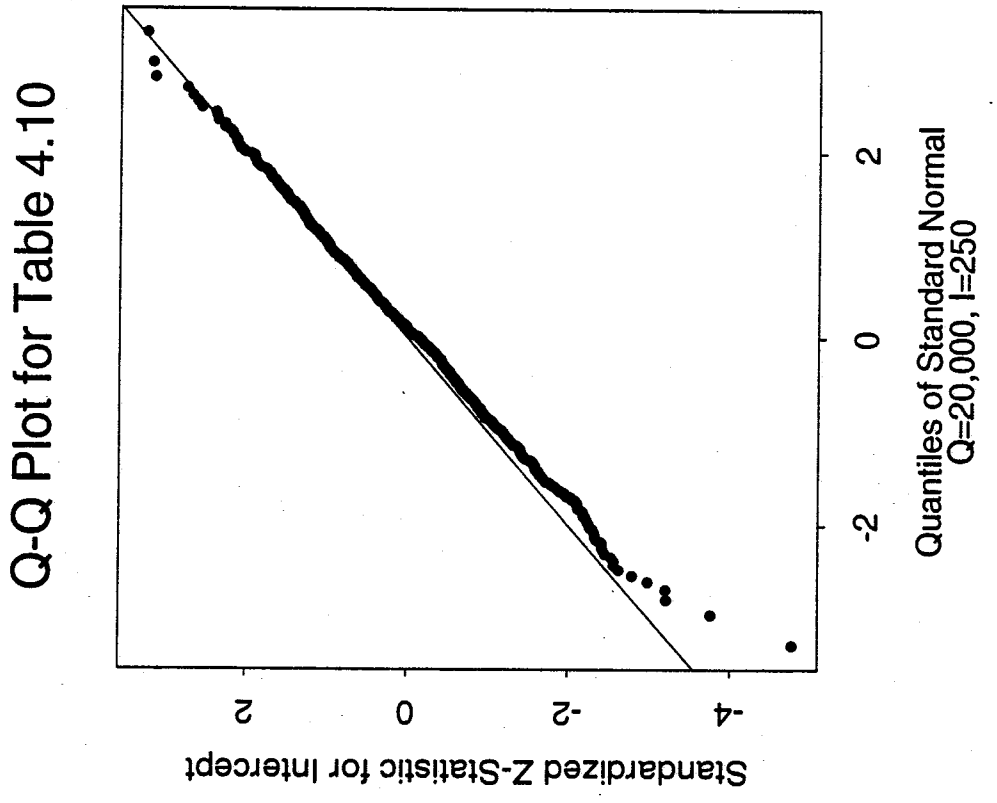

Q-Q Plot for Table 4.7
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q-Q Plot for Table 4.7
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal

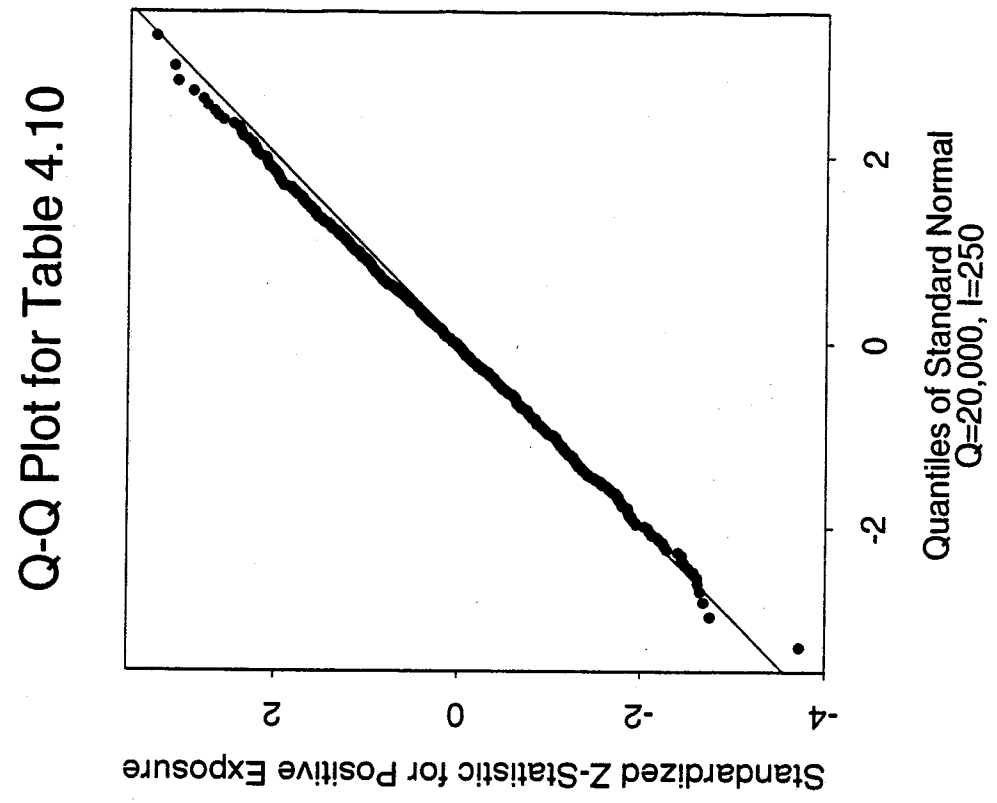

Q-Q Plot for Table 4.9
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=2000
Q-Q Plot for Table 4.9
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=2000

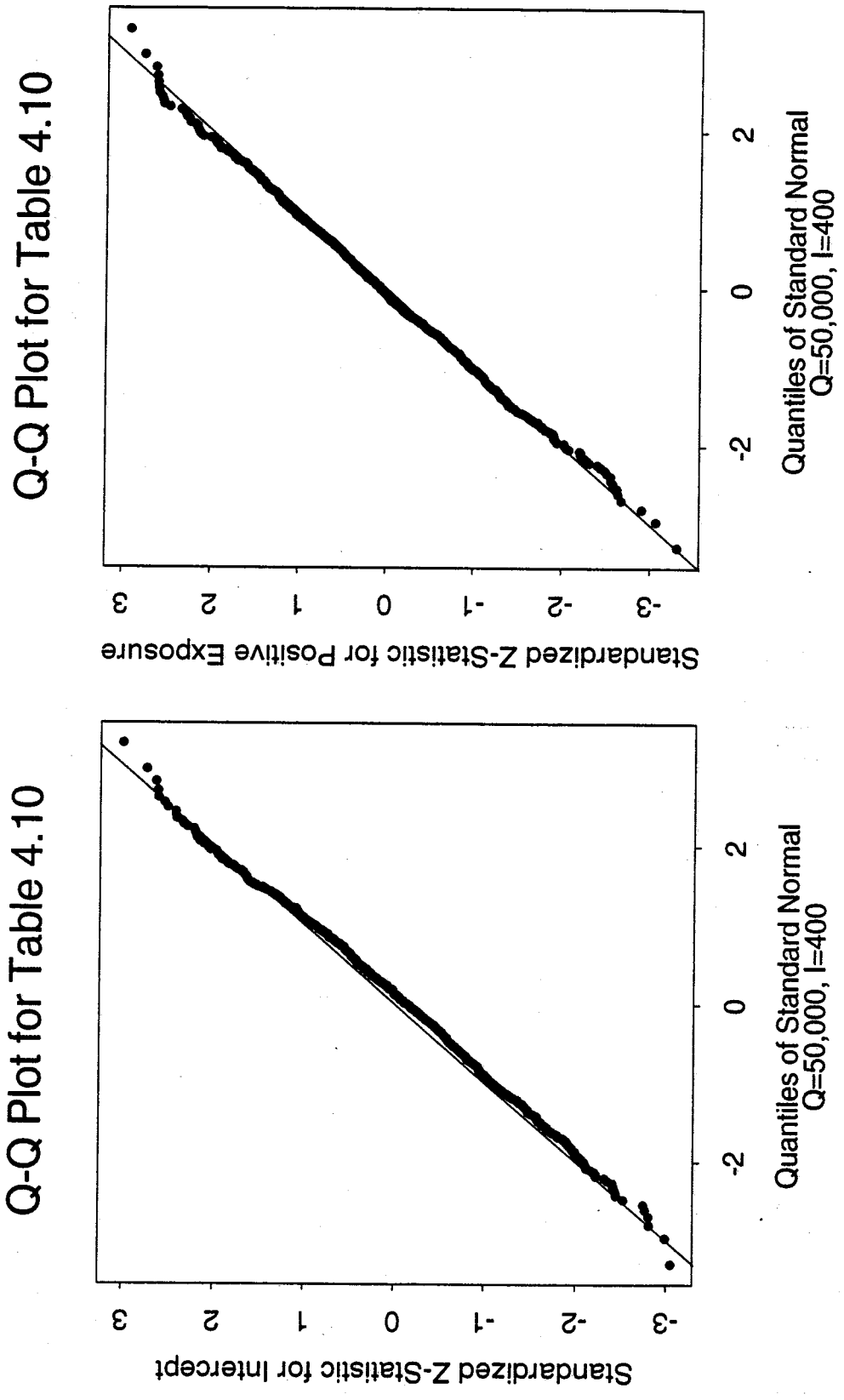

Q-Q Plot for Table 4.9
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q=20,000
-2
0
2
4
-2
0
2
Q-Q Plot for Table 4.9
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
Q=20,000
-2
0
2
4
-2
0
2

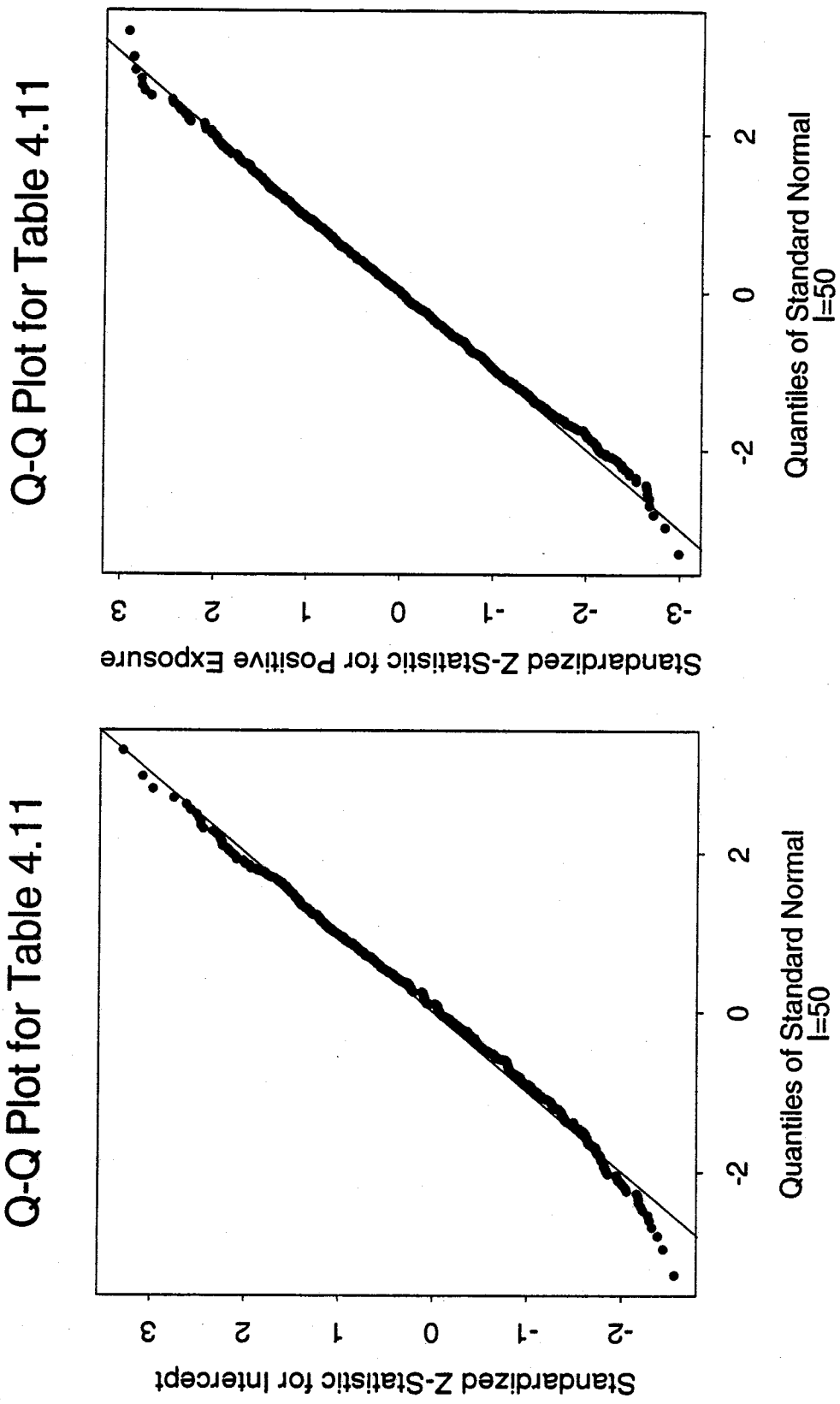

Q-Q Plot for Table 4.10
Standardized Z-Statistic for Intercept
6
4
2
0
-2
-4
-2
0
2
Quantiles of Standard Normal
Q=2000, I=250

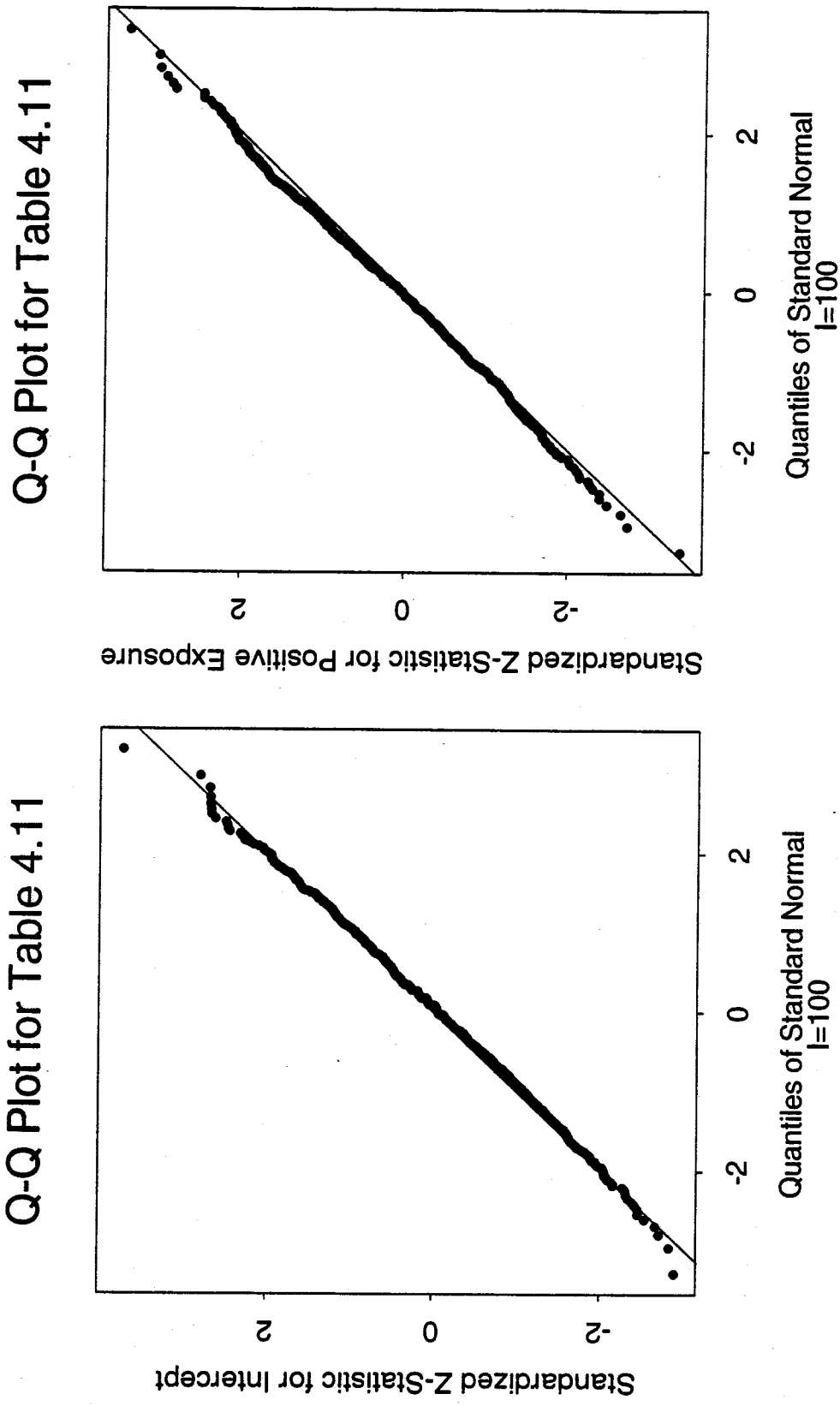

Q-Q Plot for Table 4.10
Standardized Z-Statistic for Positive Exposure
4
2
0
-2
-2
0
2
Quantiles of Standard Normal
Q=2000, I=250

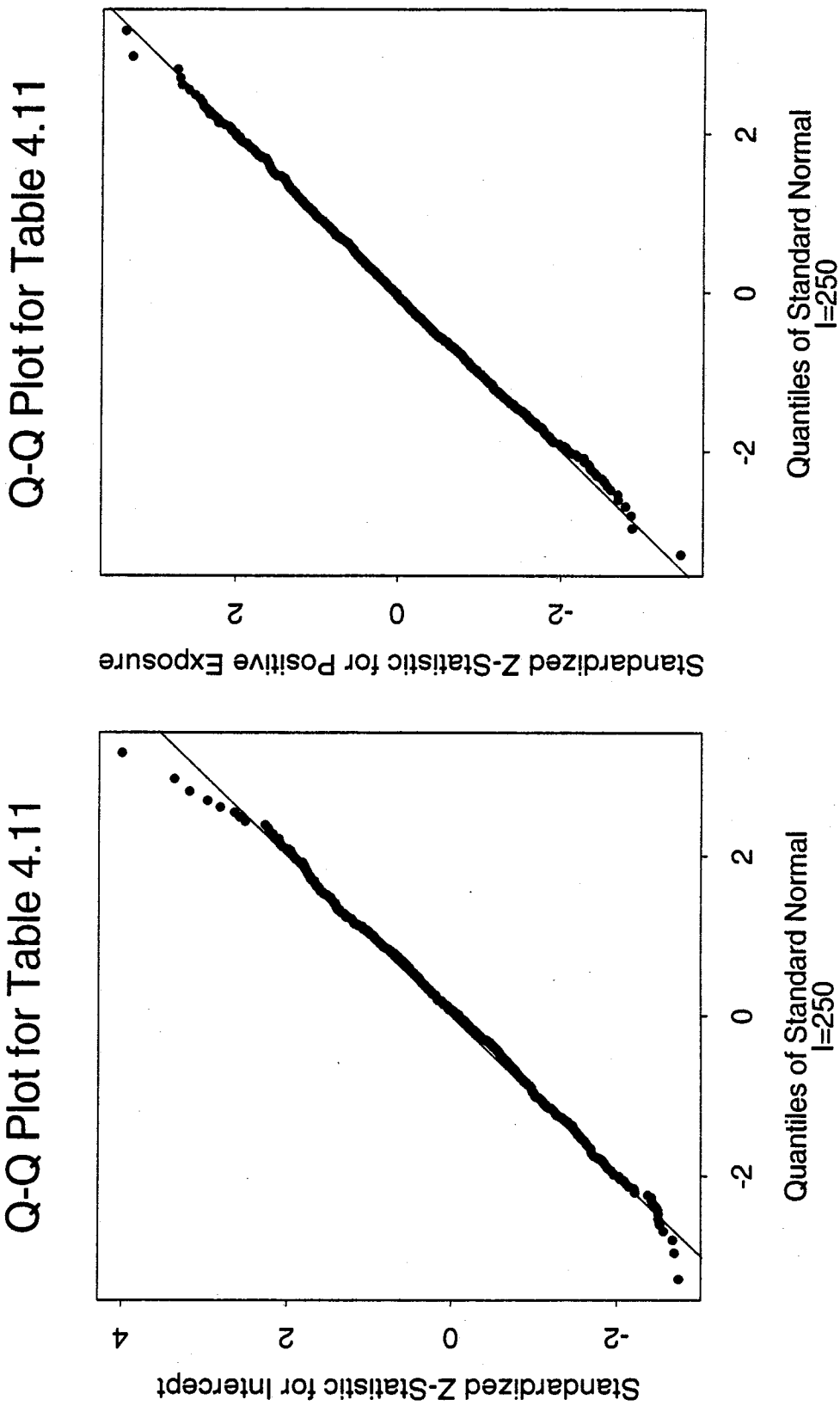

Q-Q Plot for Table 4.10
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=20,000, I=250

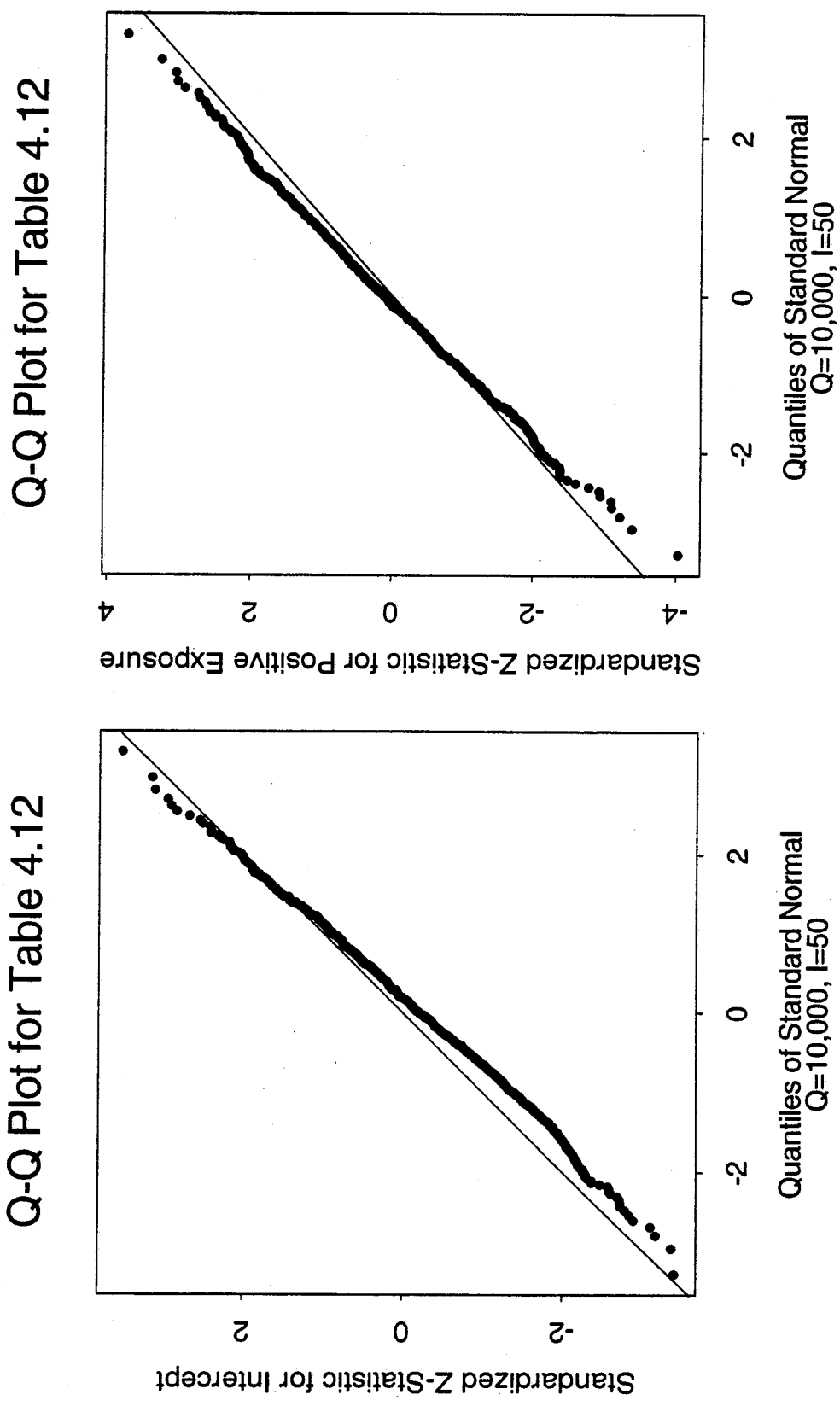

Q-Q Plot for Table 4.10
Standardized Z-Statistic for Positive Exposure
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=20,000, I=250

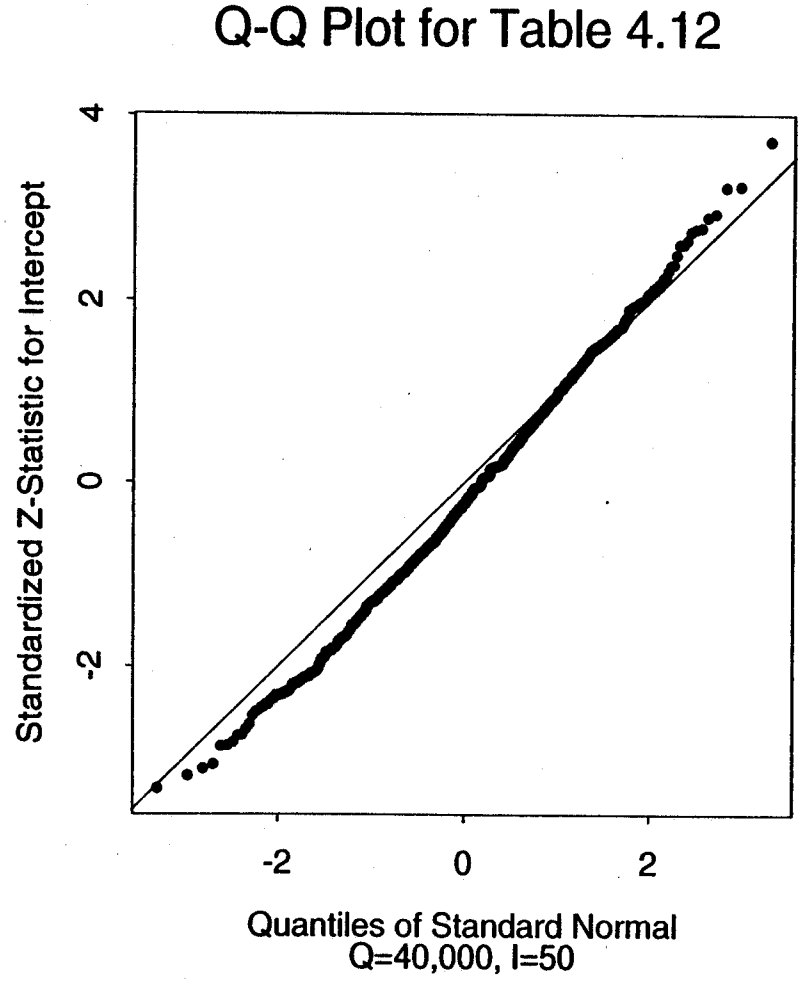

Q-Q Plot for Table 4.10
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=50,000, I=400
Q-Q Plot for Table 4.10
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=50,000, I=400

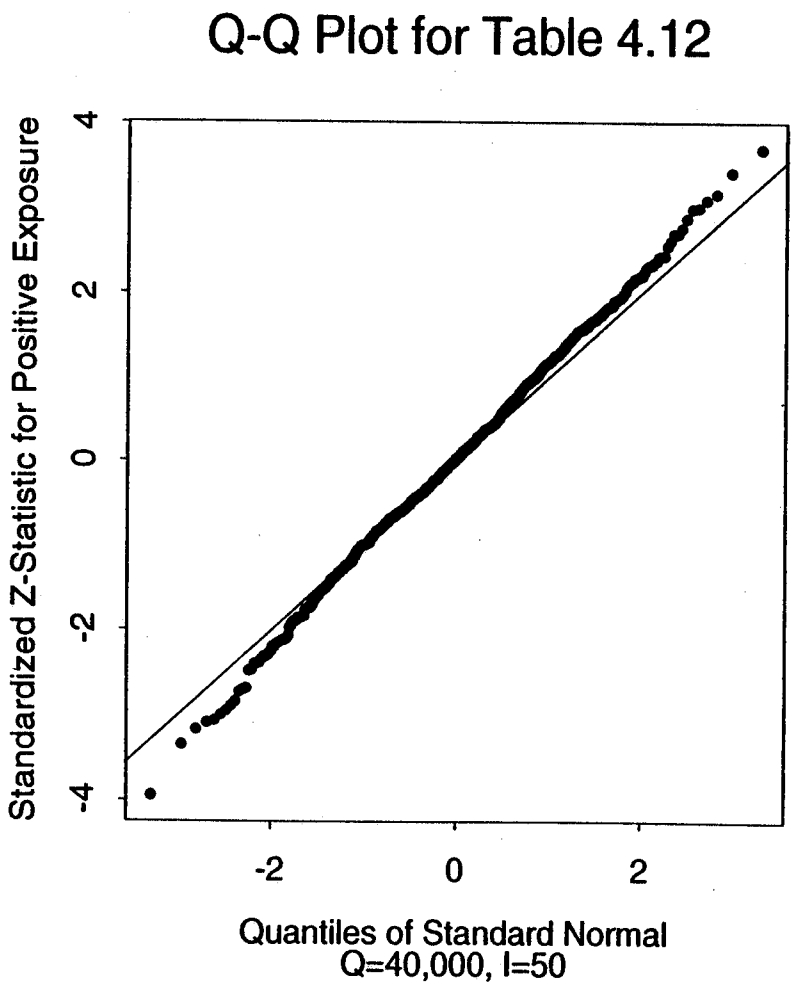

Q-Q Plot for Table 4.11
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=50
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.11
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=50
-3
-2
-1
0
1
2
3
-2
0
2

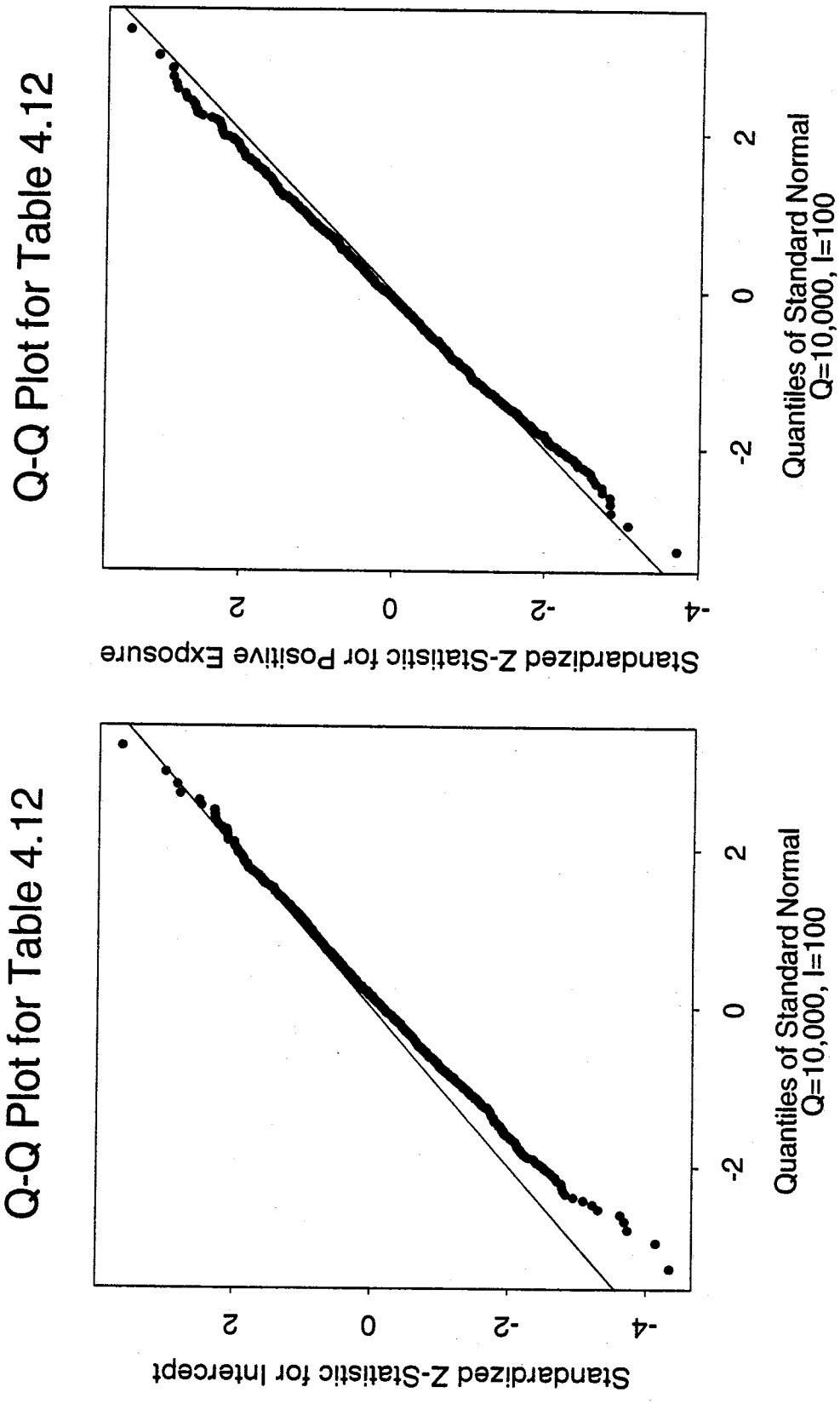

Q-Q Plot for Table 4.11
Standardized Z-Statistic for Intercept
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=100
Q-Q Plot for Table 4.11
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=100

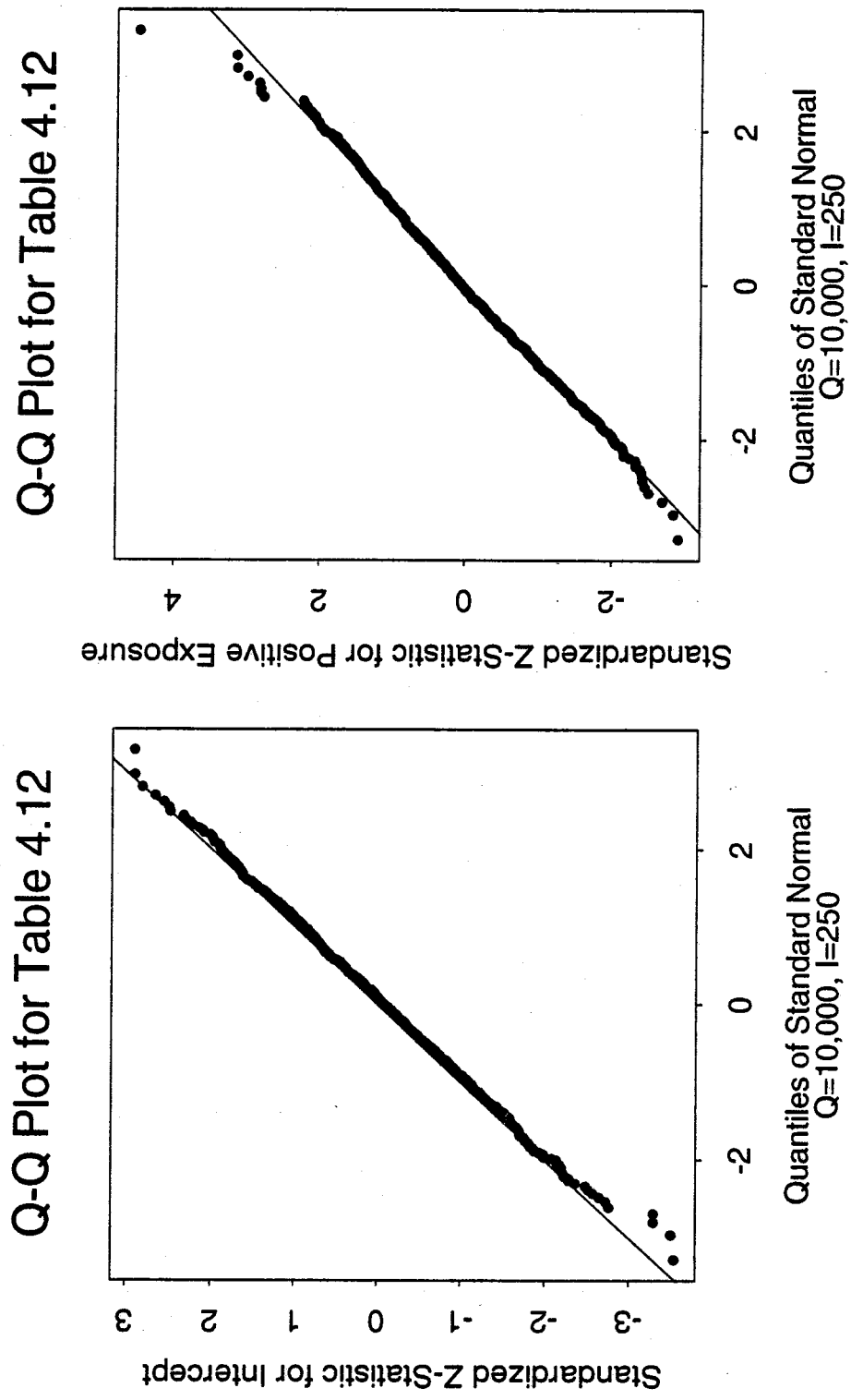

Q-Q Plot for Table 4.11
Standardized Z-Statistic for Intercept
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
I=250
Q-Q Plot for Table 4.11
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=250

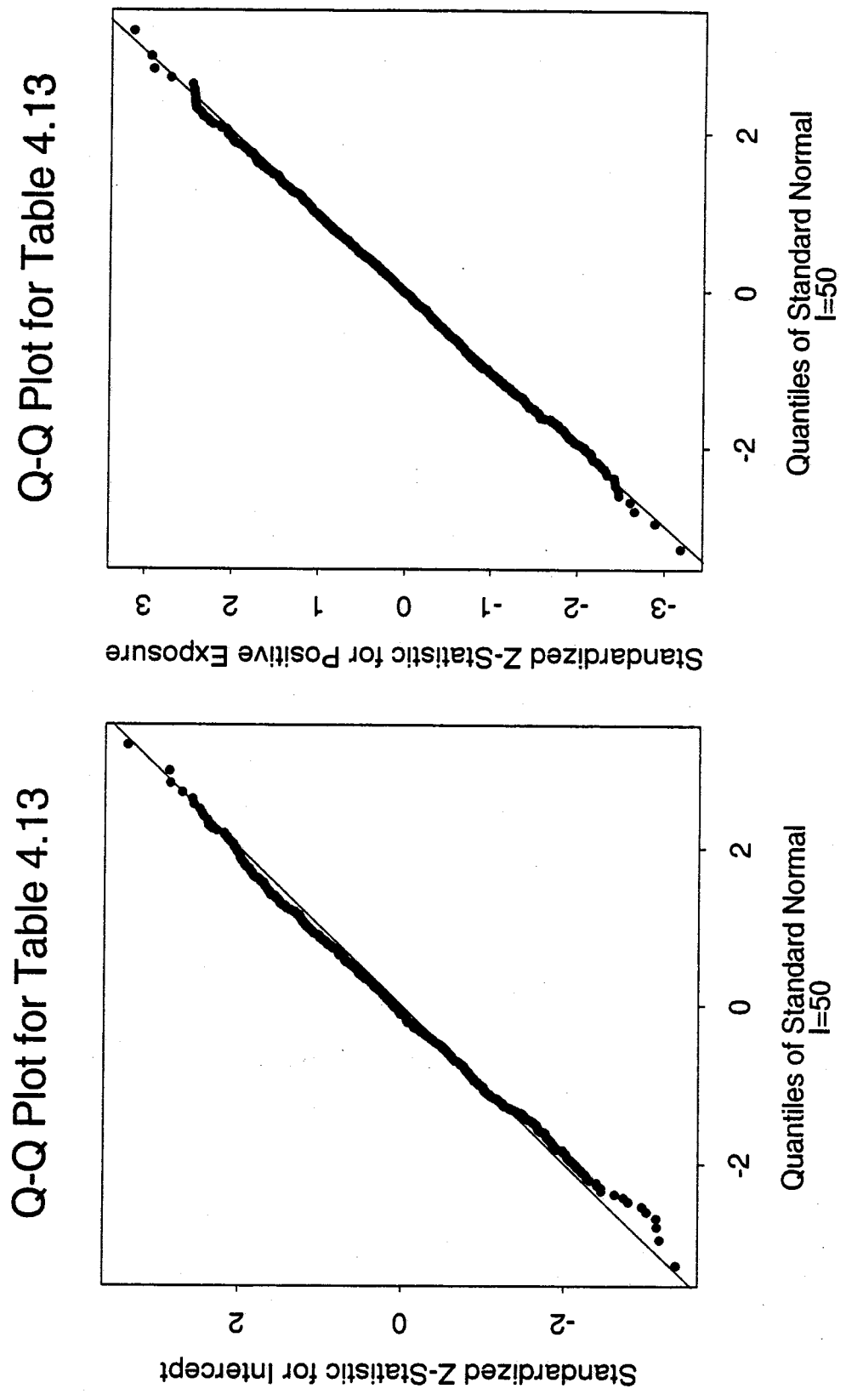

Q-Q Plot for Table 4.12
Standardized Z-Statistic for Intercept
2
0
-2
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=50
Q-Q Plot for Table 4.12
Standardized Z-Statistic for Positive Exposure
4
2
0
-2
-4
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=50

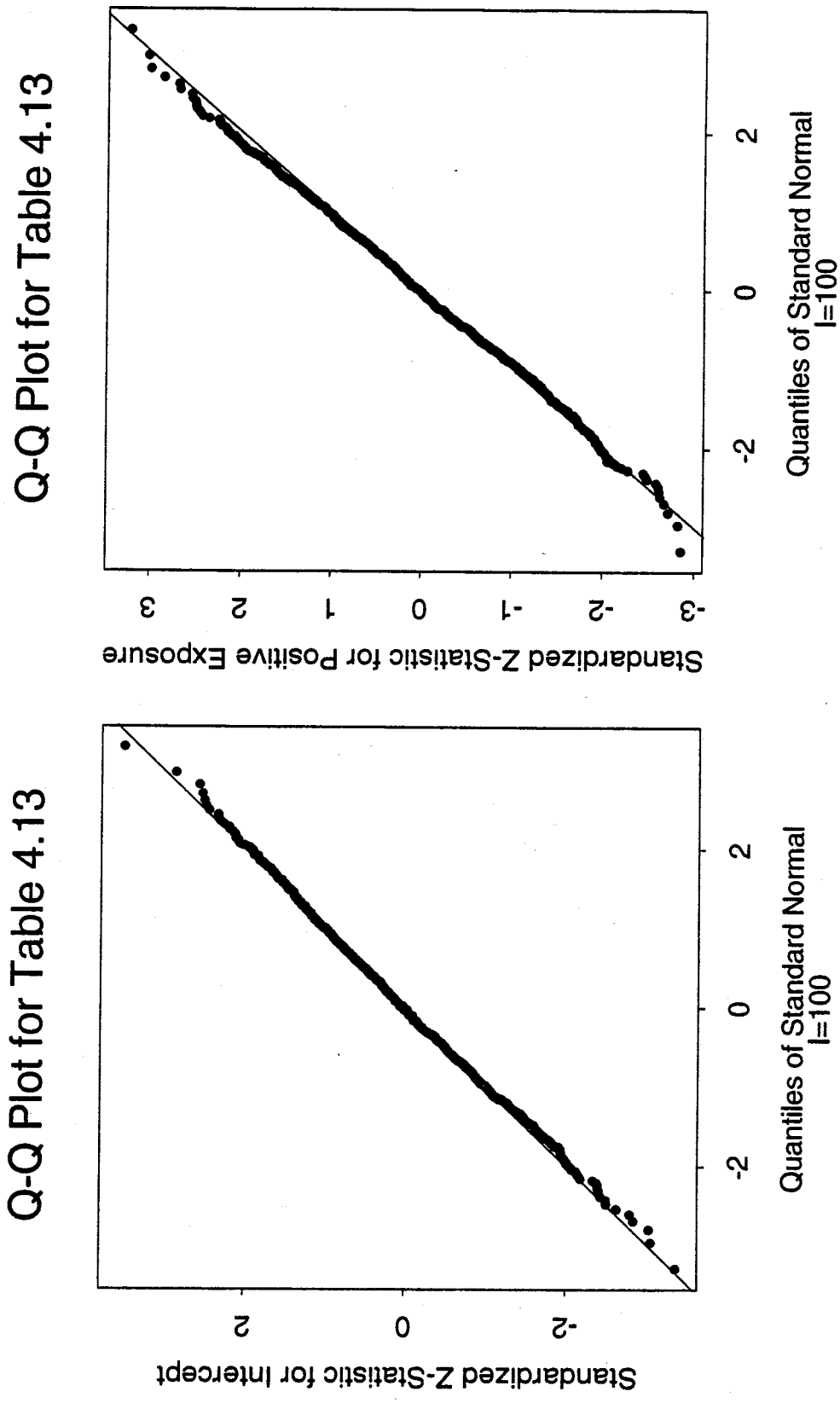

Q-Q Plot for Table 4.12
Standardized Z-Statistic for Intercept
4
2
0
-2
-2
0
2
Quantiles of Standard Normal
Q=40,000, I=50

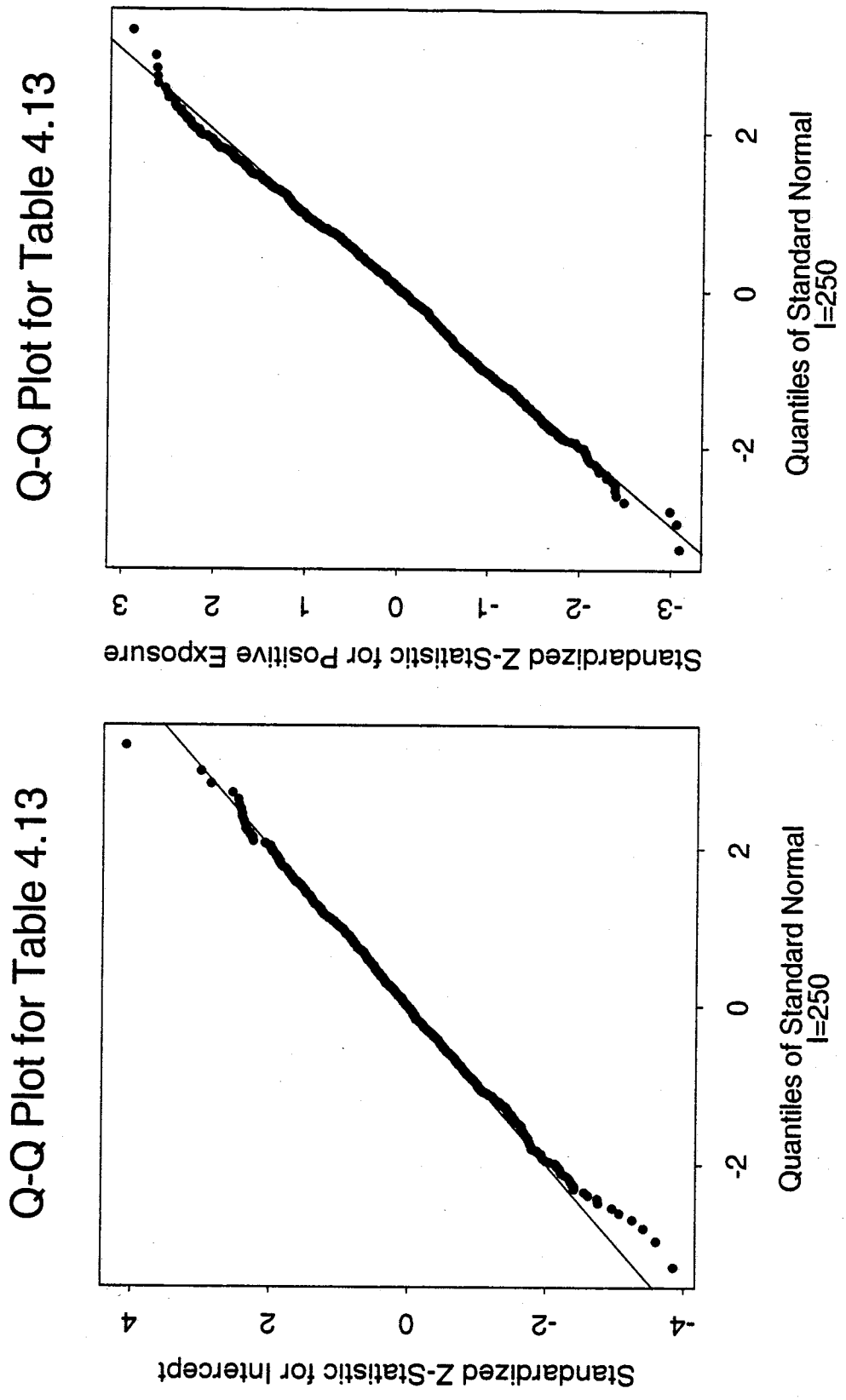

Q-Q Plot for Table 4.12
Standardized Z-Statistic for Positive Exposure
4
2
0
-2
-4
-2
0
2
Quantiles of Standard Normal
Q=40,000, I=50

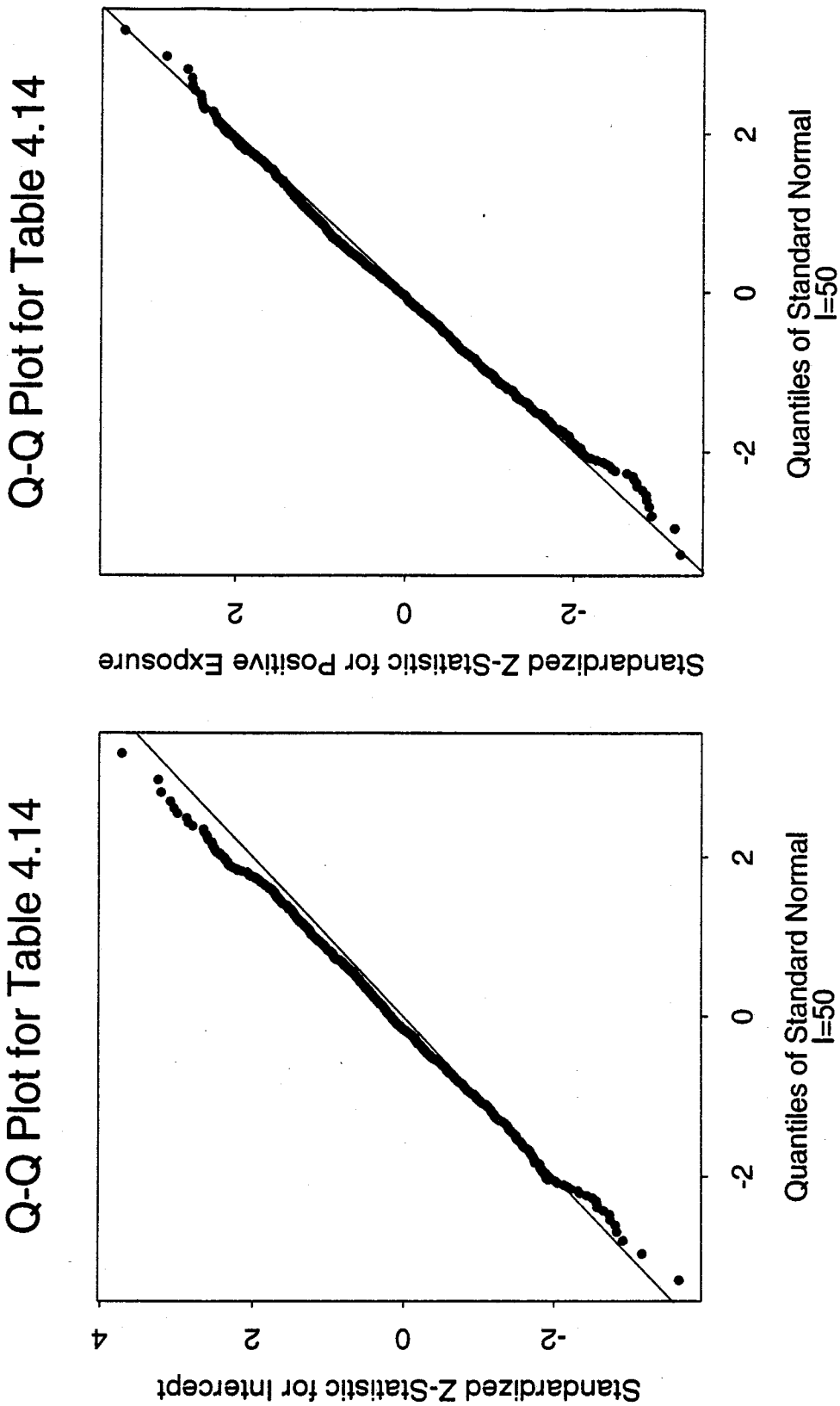

Q-Q Plot for Table 4.12
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=100
Q-Q Plot for Table 4.12
Standardized Z-Statistic for Positive Exposure
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=100

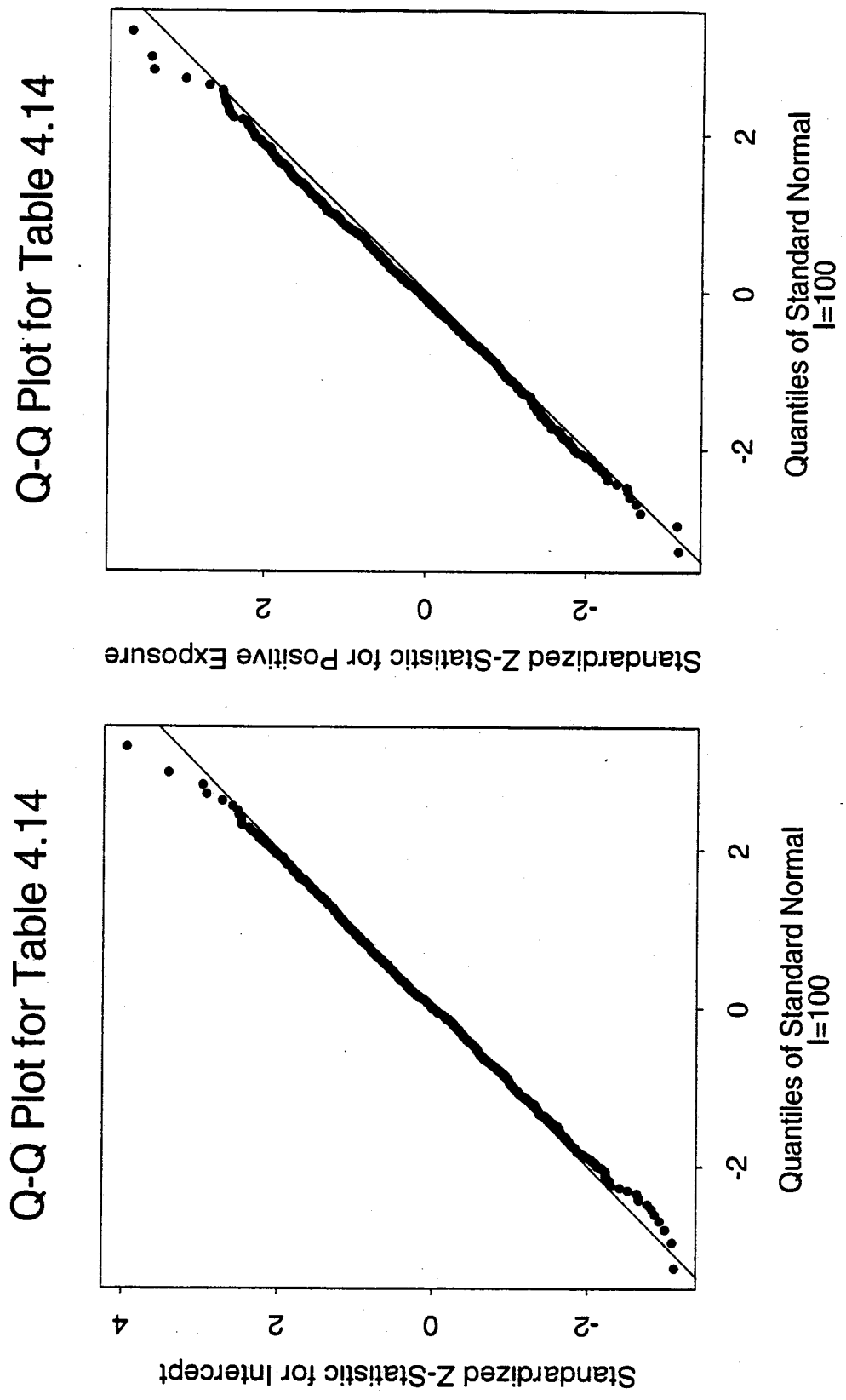

Q-Q Plot for Table 4.12
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=250
Q-Q Plot for Table 4.12
Standardized Z-Statistic for Positive Exposure
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
Q=10,000, I=250

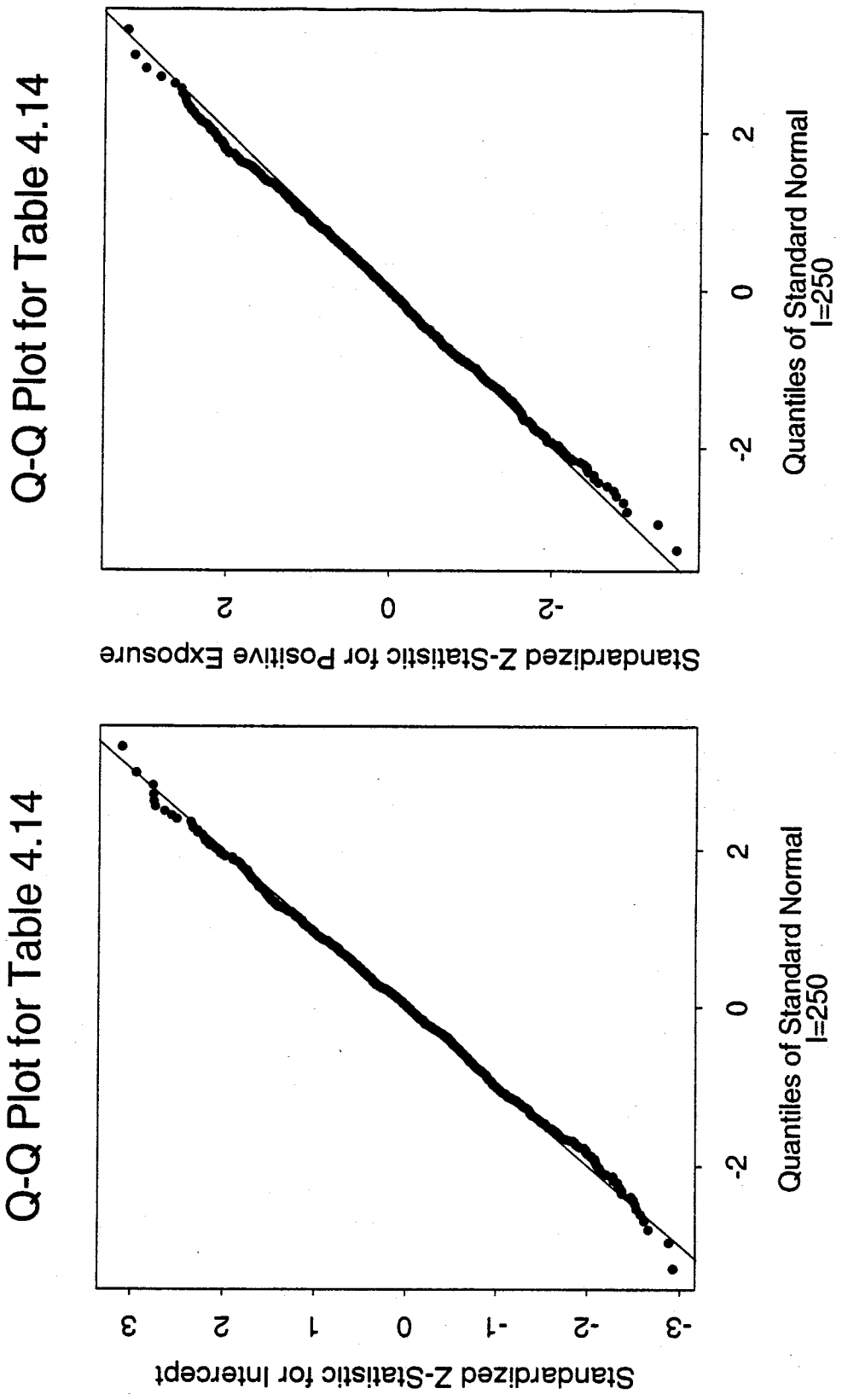

Q-Q Plot for Table 4.13
Standardized Z-Statistic for Intercept
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=50
Q-Q Plot for Table 4.13
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=50

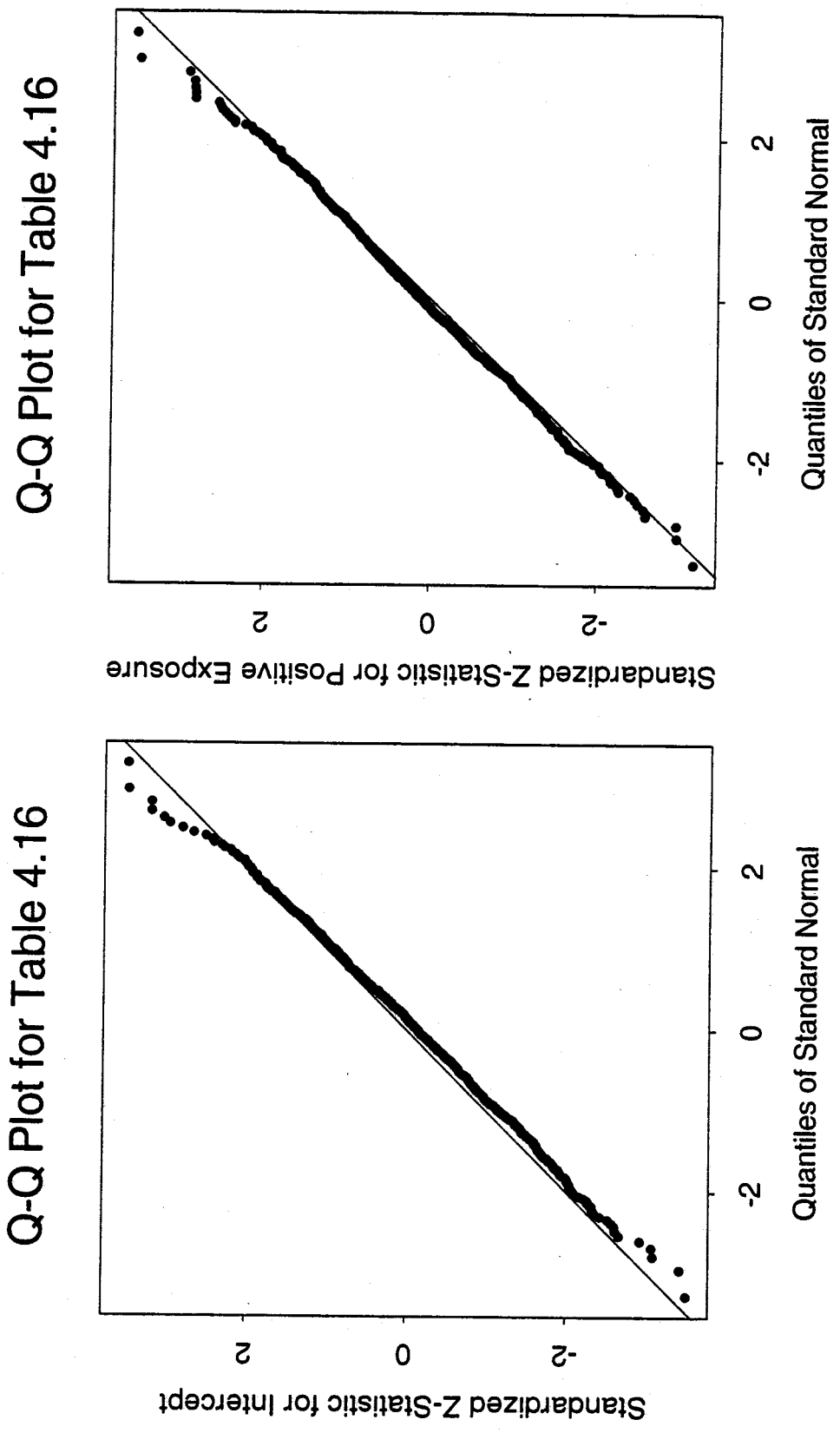

Q-Q Plot for Table 4.13
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=100
-3
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.13
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=100
-2
0
2
-2
0
2

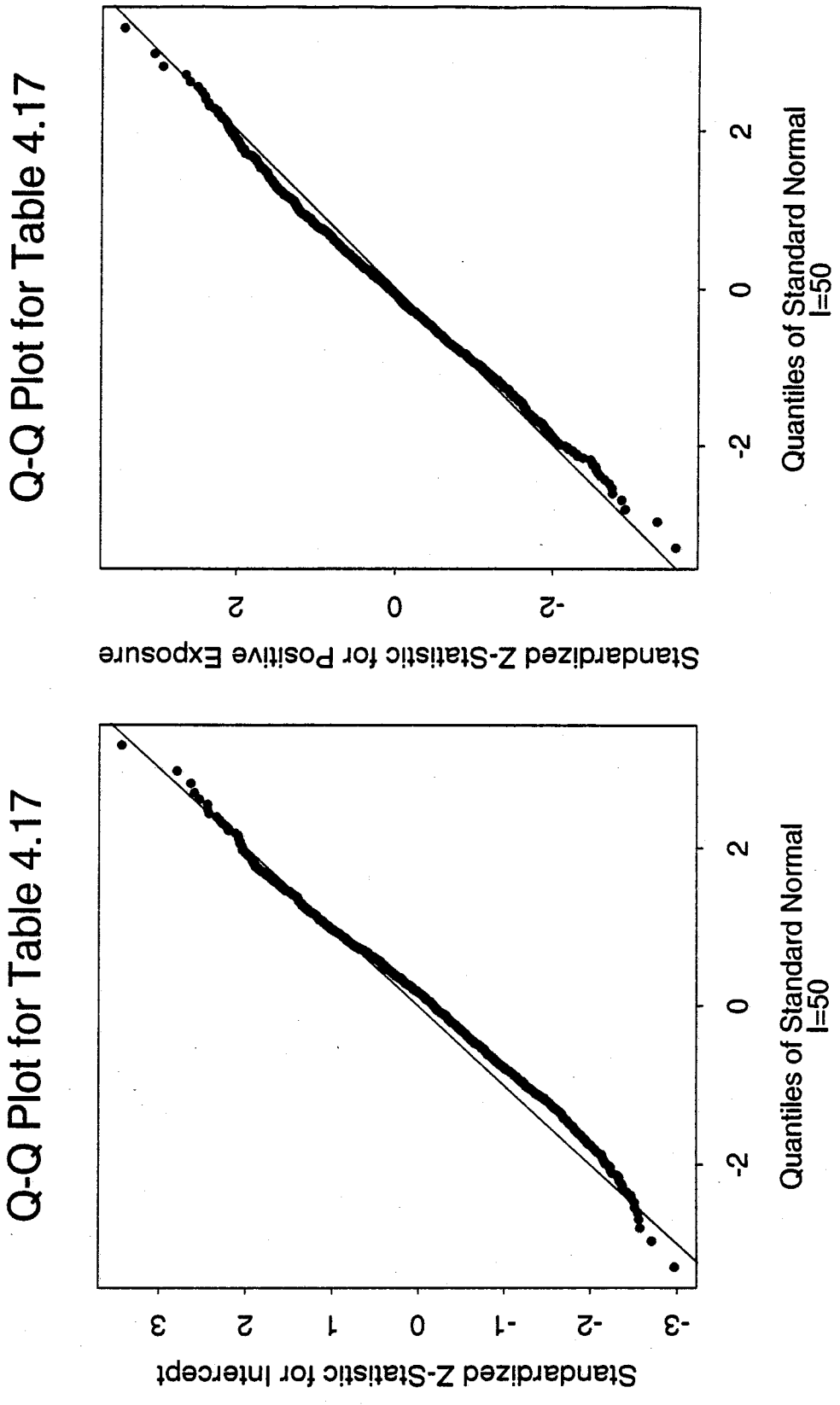

Q-Q Plot for Table 4.13
Standardized Z-Statistic for Intercept
-4
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
I=250
Q-Q Plot for Table 4.13
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250

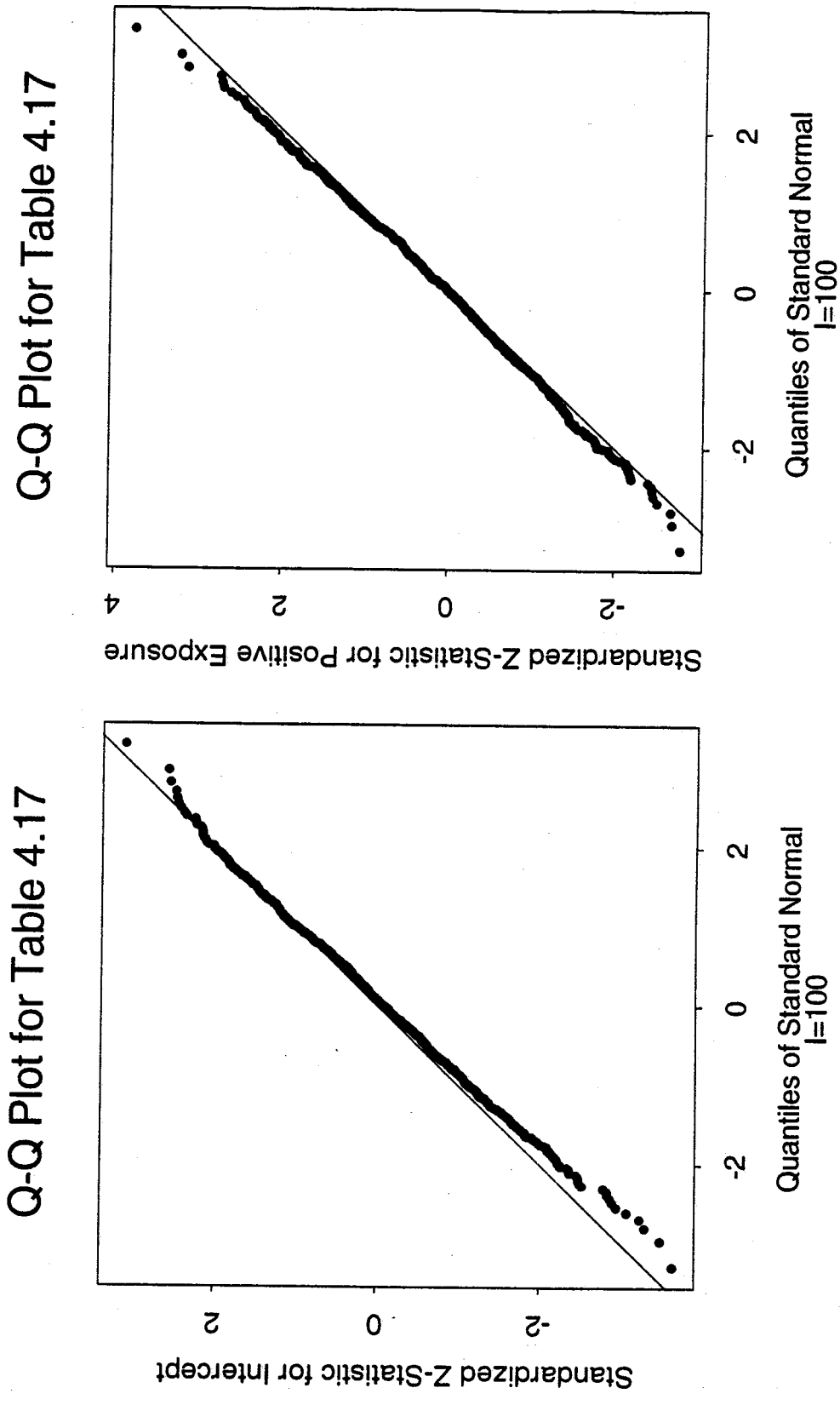

Q-Q Plot for Table 4.14
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=50
Q-Q Plot for Table 4.14
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=50

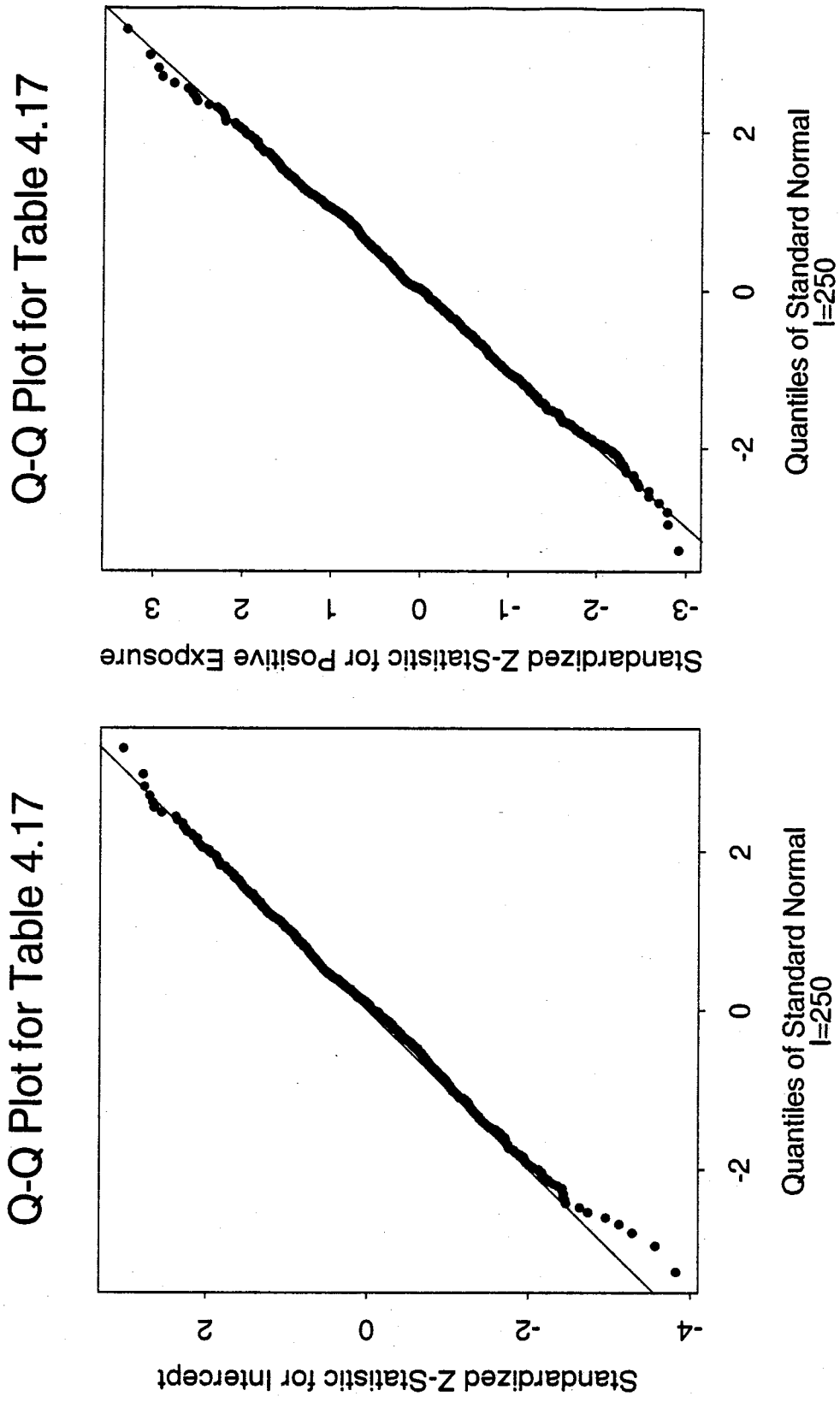

Q-Q Plot for Table 4.14
Standardized Z-Statistic for Intercept
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
I=100
Q-Q Plot for Table 4.14
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=100

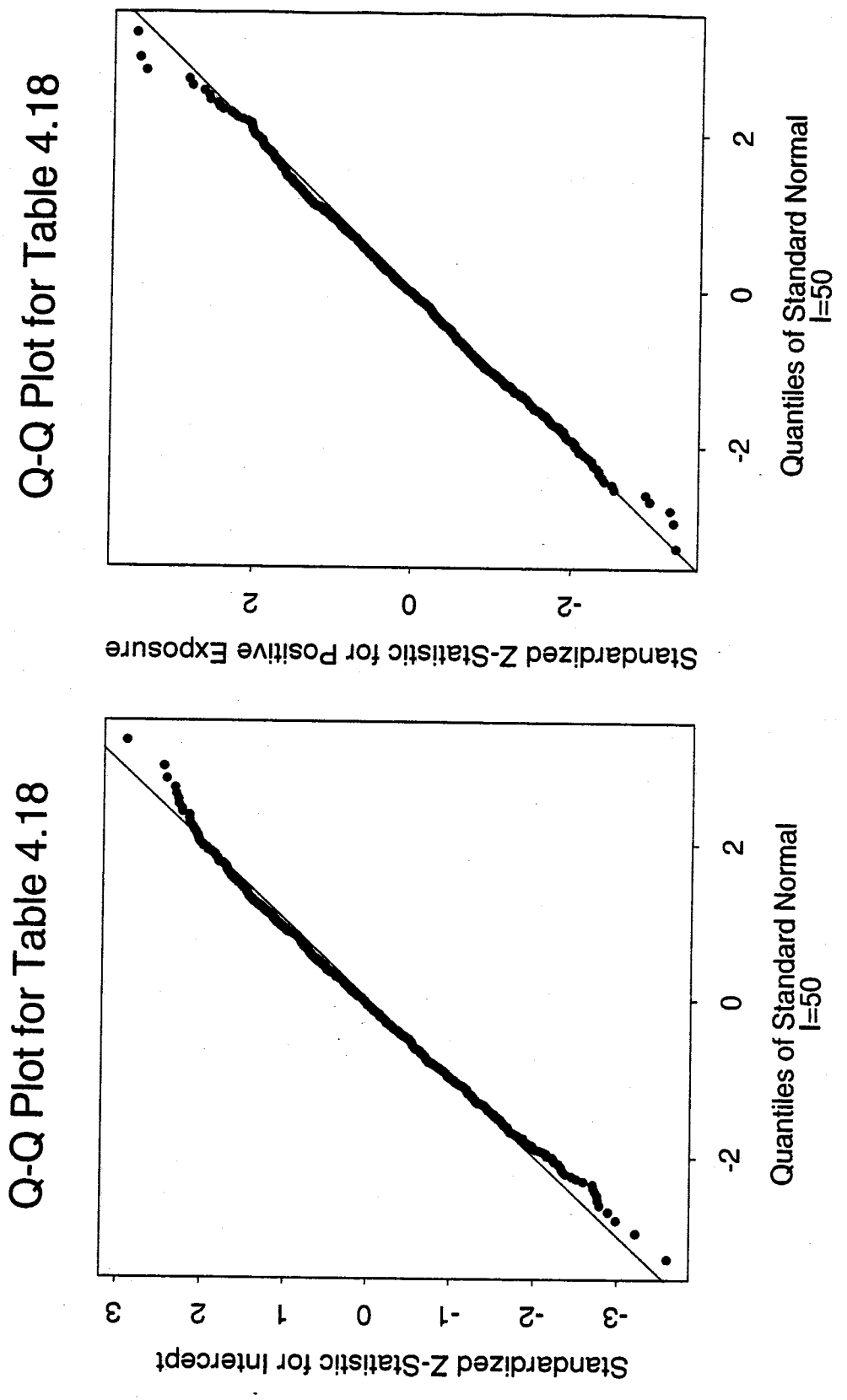

Q-Q Plot for Table 4.14
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250
Q-Q Plot for Table 4.14
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=250

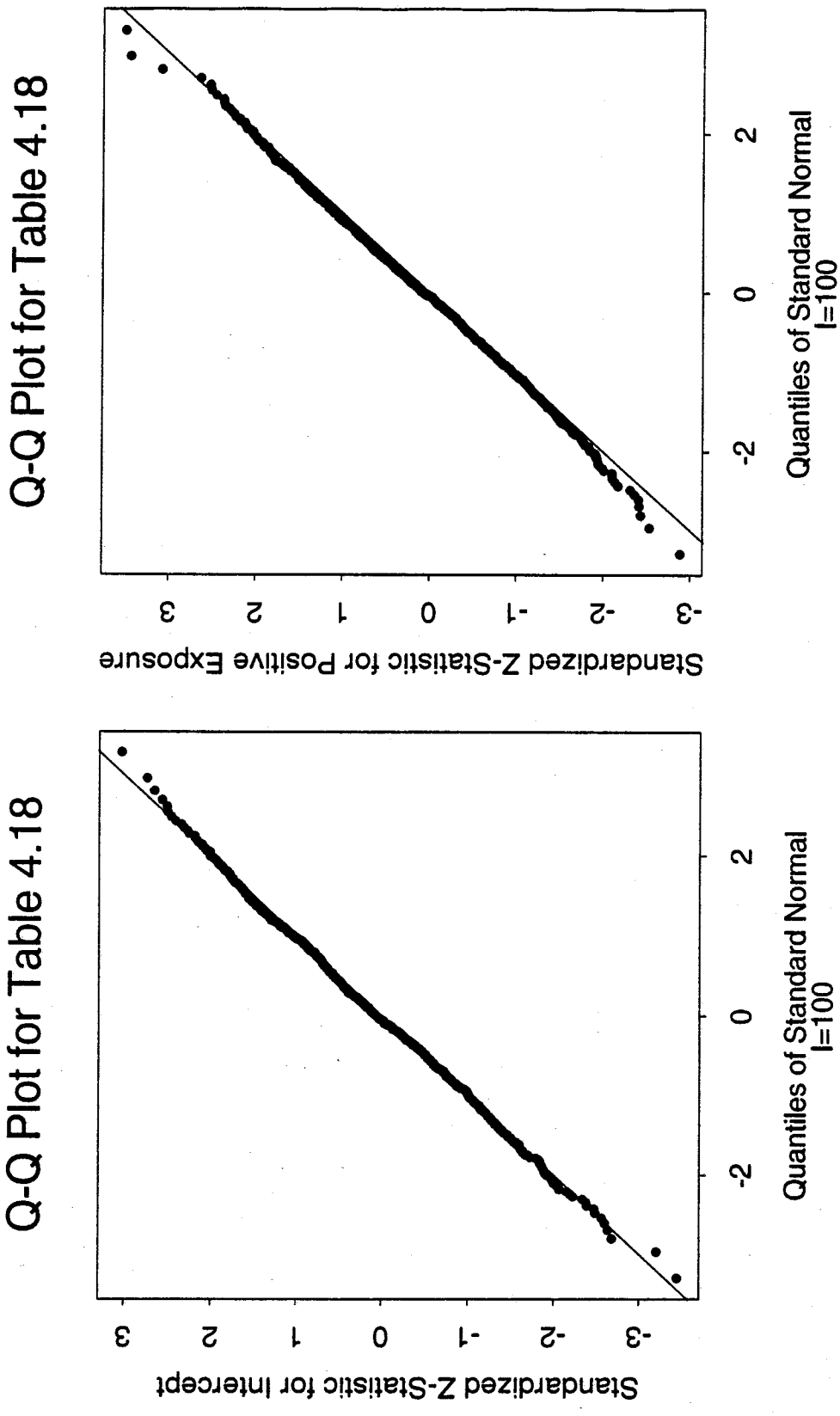

Q-Q Plot for Table 4.16
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
-2
0
2
Q-Q Plot for Table 4.16
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
-2
0
2

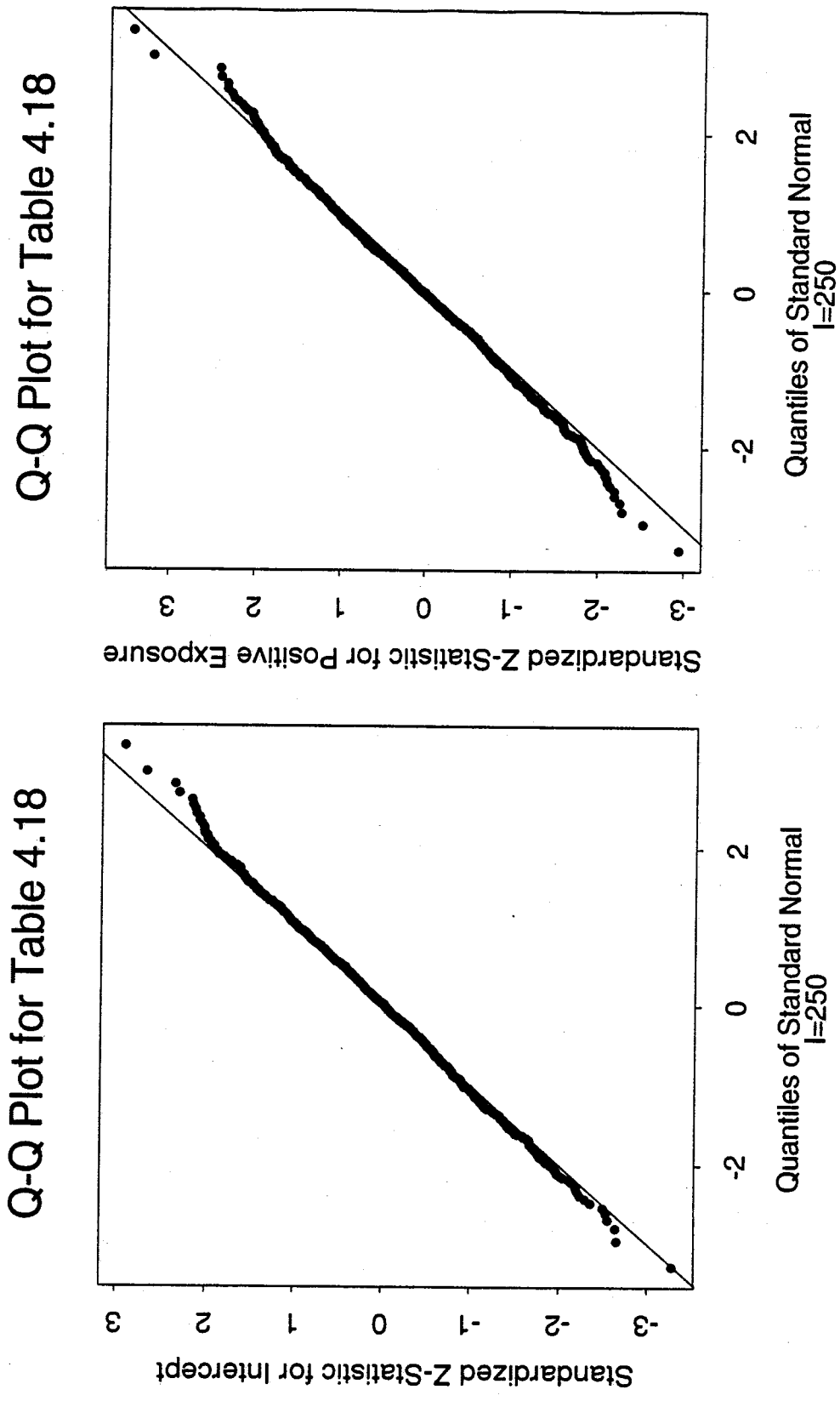

Q-Q Plot for Table 4.17
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=50
Q-Q Plot for Table 4.17
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=50

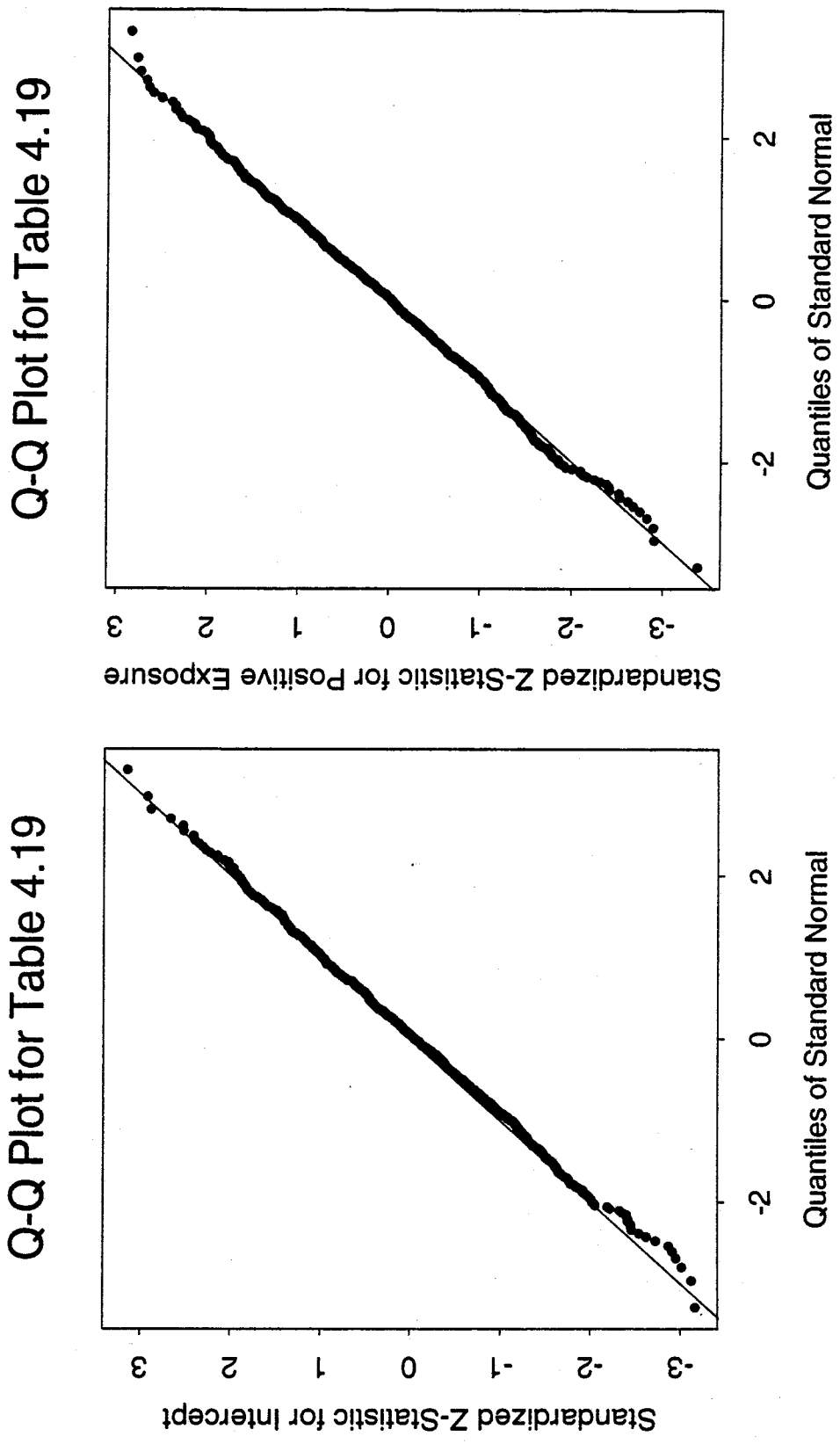

Q-Q Plot for Table 4.17
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=100
-2
0
2
Q-Q Plot for Table 4.17
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=100
-2
0
2
4

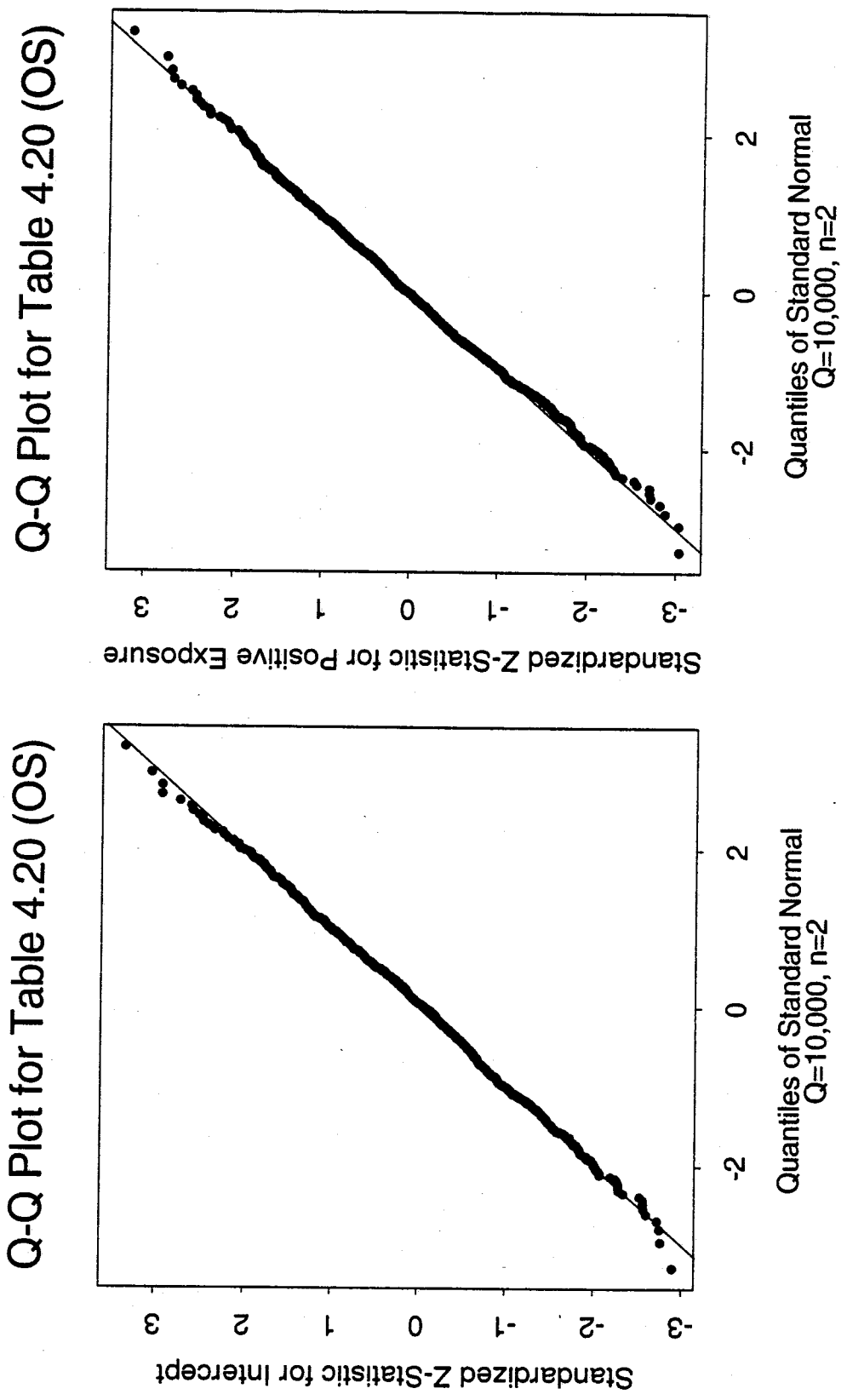

Q-Q Plot for Table 4.17
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=250
Q-Q Plot for Table 4.17
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250

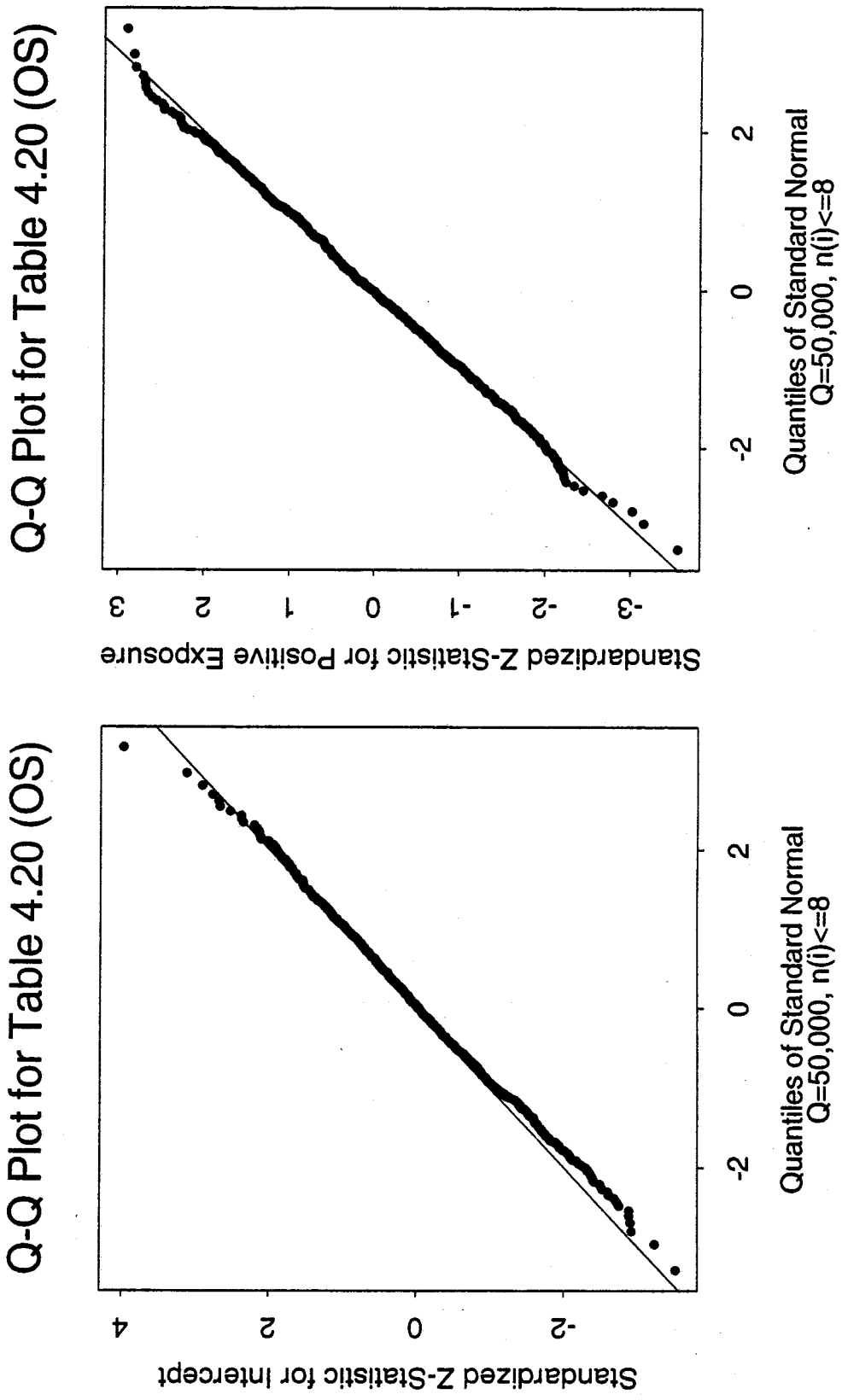

Q-Q Plot for Table 4.18
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=50
Q-Q Plot for Table 4.18
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=50

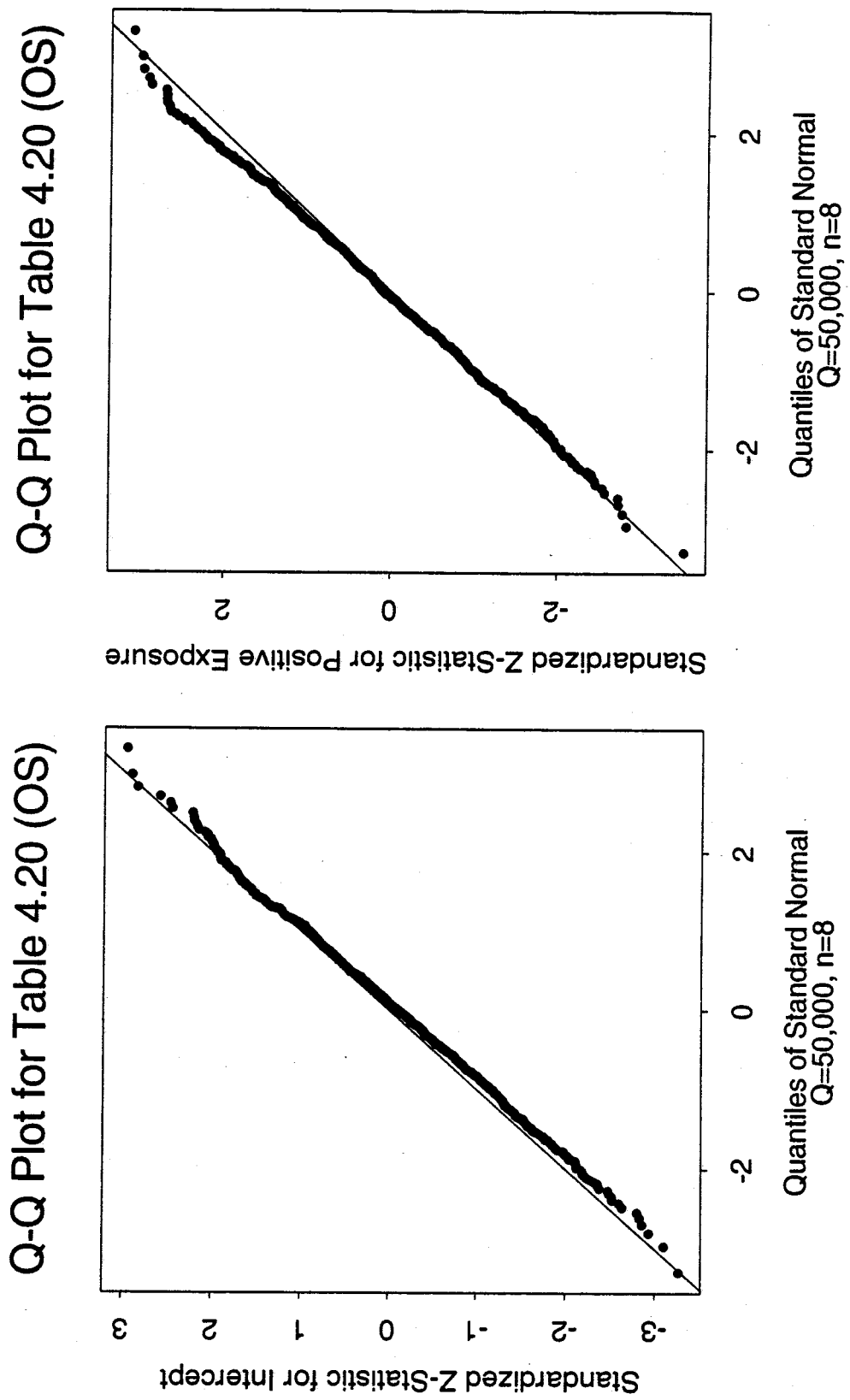

Q-Q Plot for Table 4.18
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=100
Q-Q Plot for Table 4.18
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=100

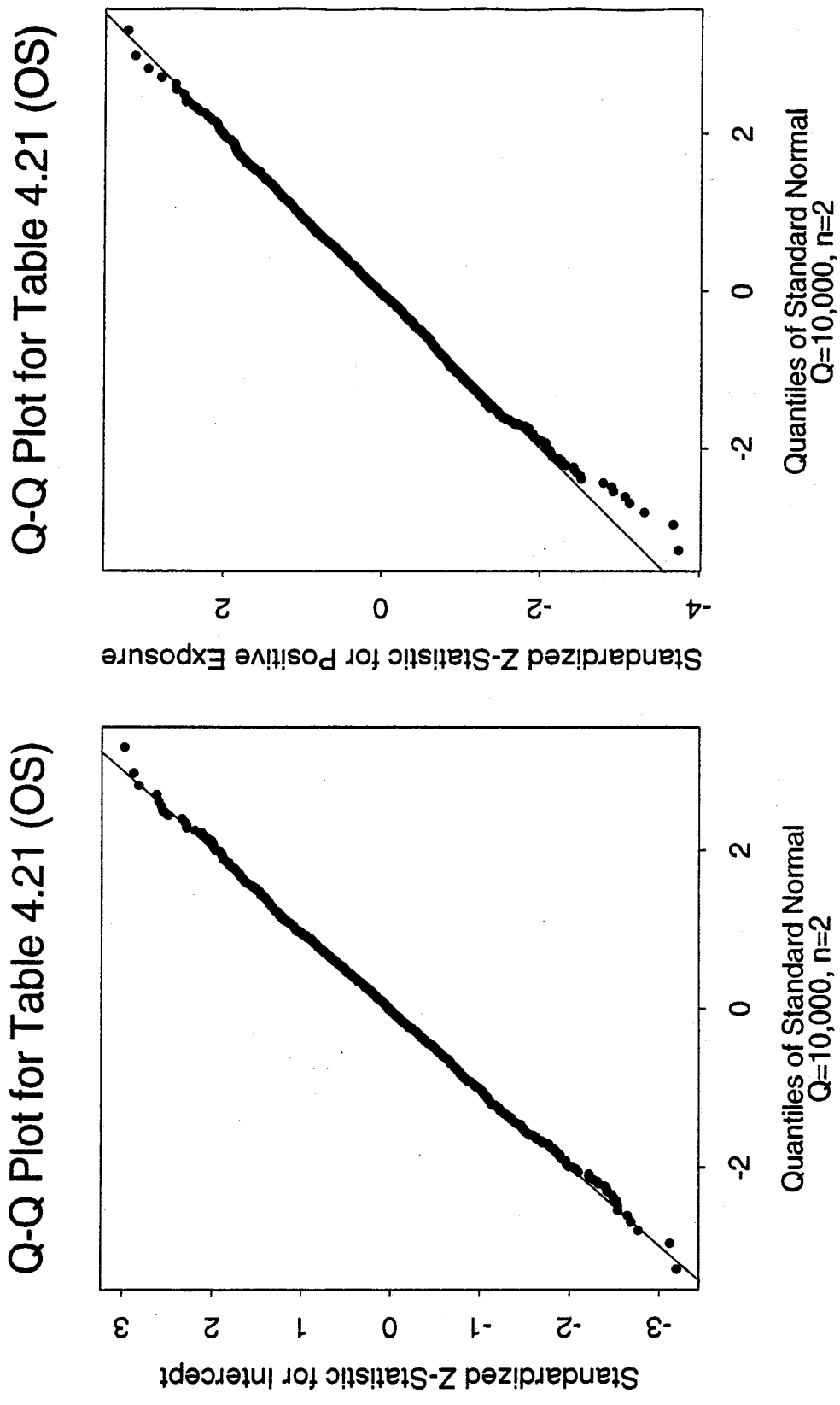

Q-Q Plot for Table 4.18
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250
Q-Q Plot for Table 4.18
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250

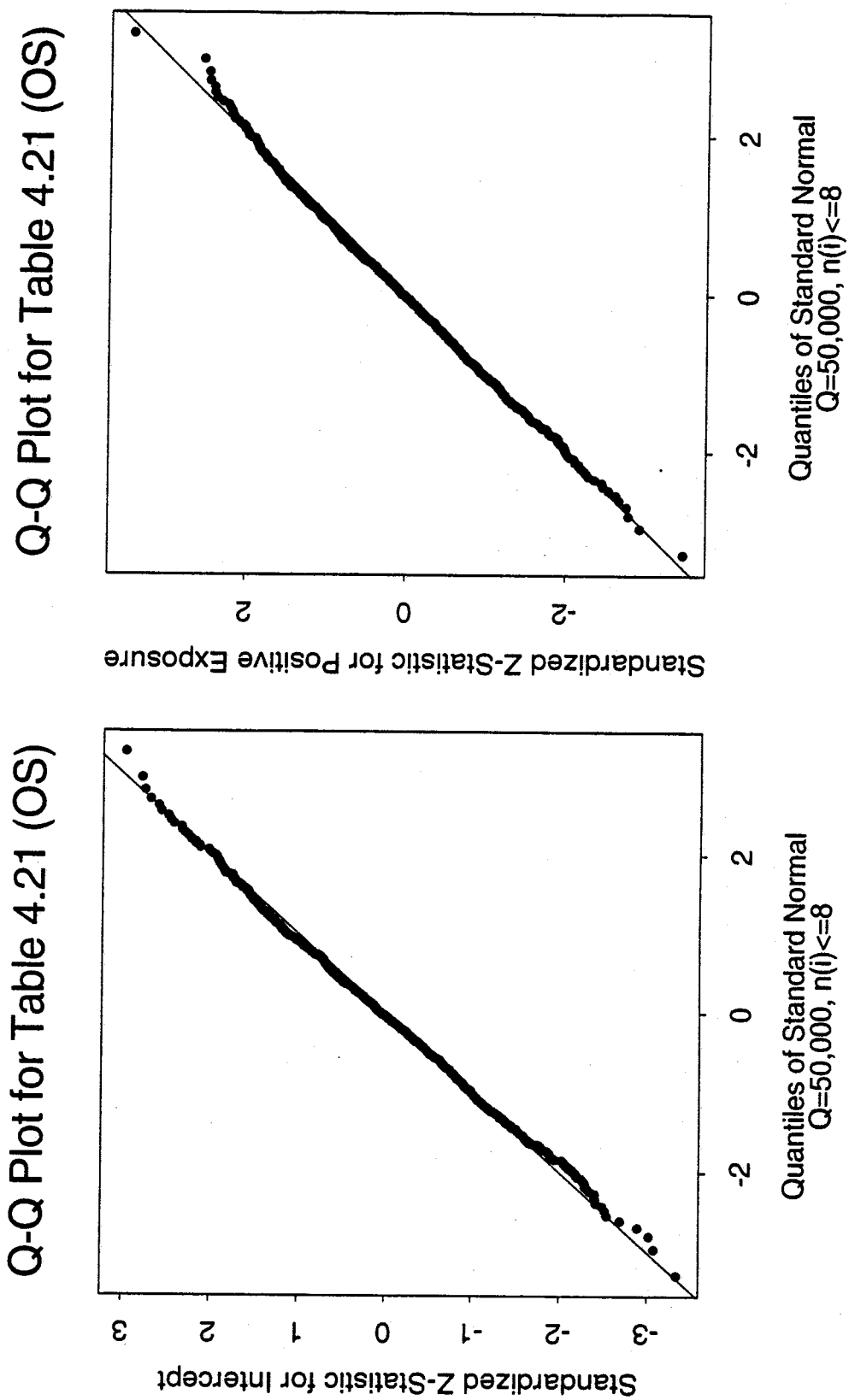

Q-Q Plot for Table 4.19
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.19
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3
-2
0
2

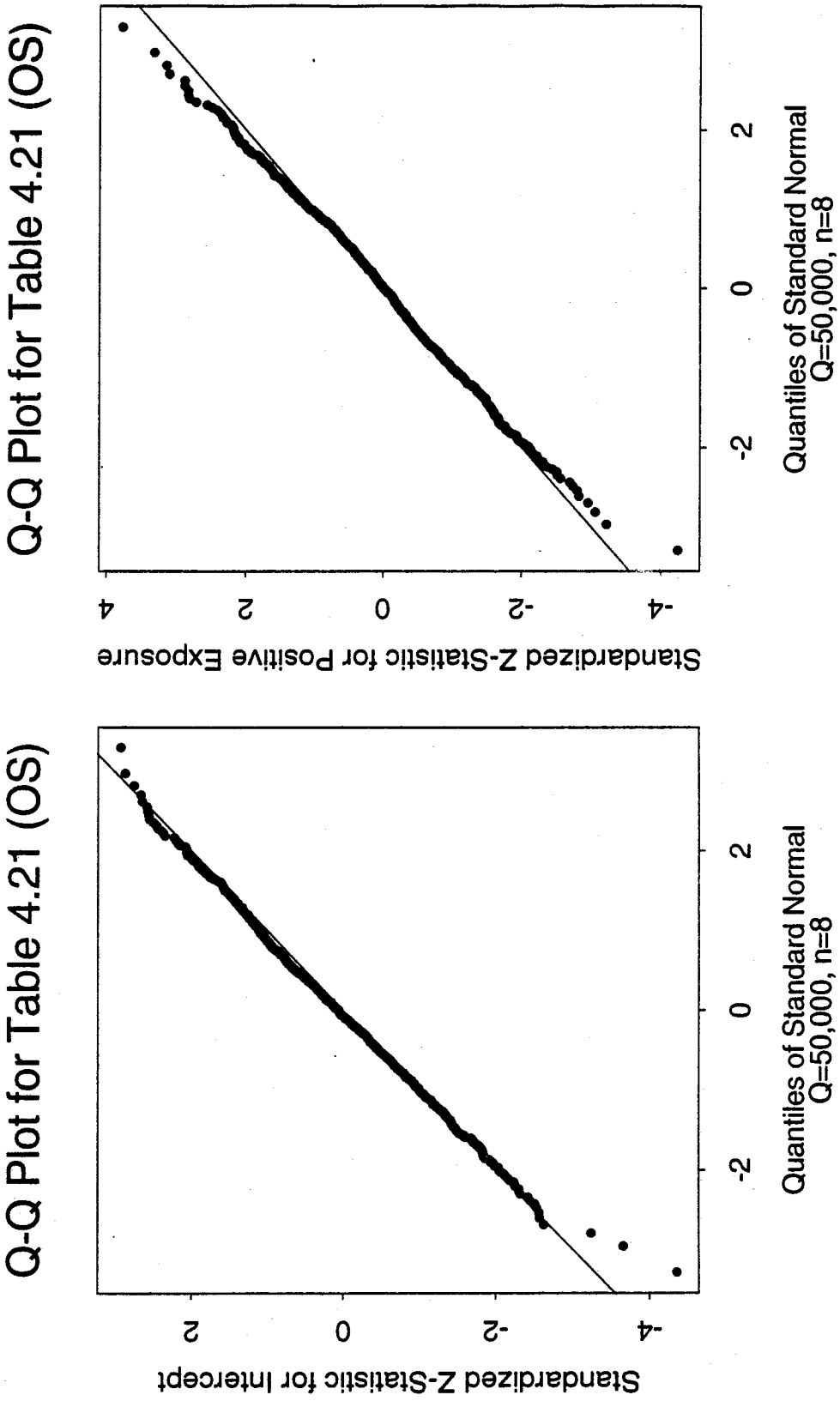

Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q=10,000, n=2
-3
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
Q=10,000, n=2

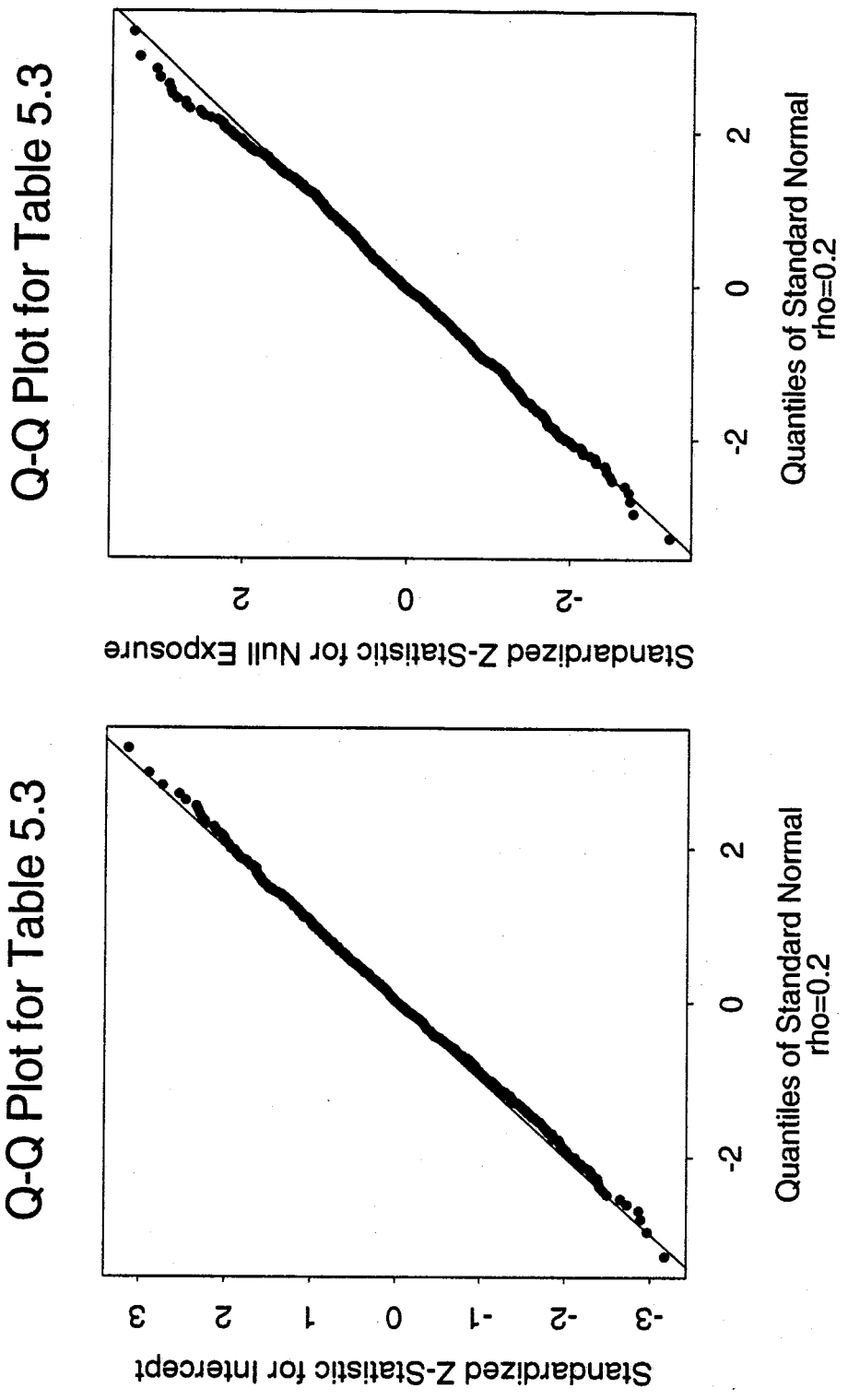

Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Intercept
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
Q=50,000, n(i)<=8
Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=50,000, n(i)<=8

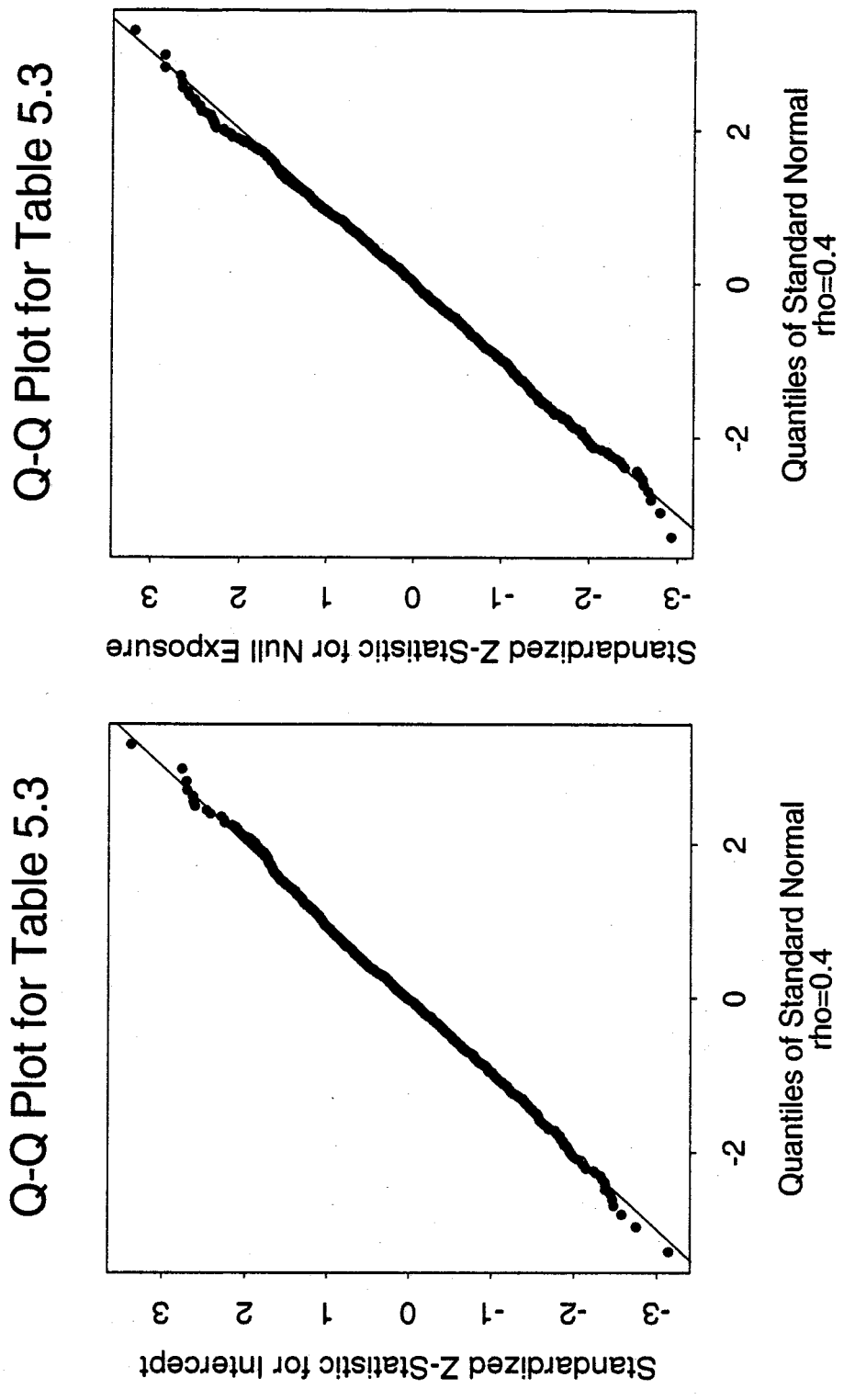

Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
Q=50,000, n=8
Q-Q Plot for Table 4.20 (OS)
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
Q=50,000, n=8

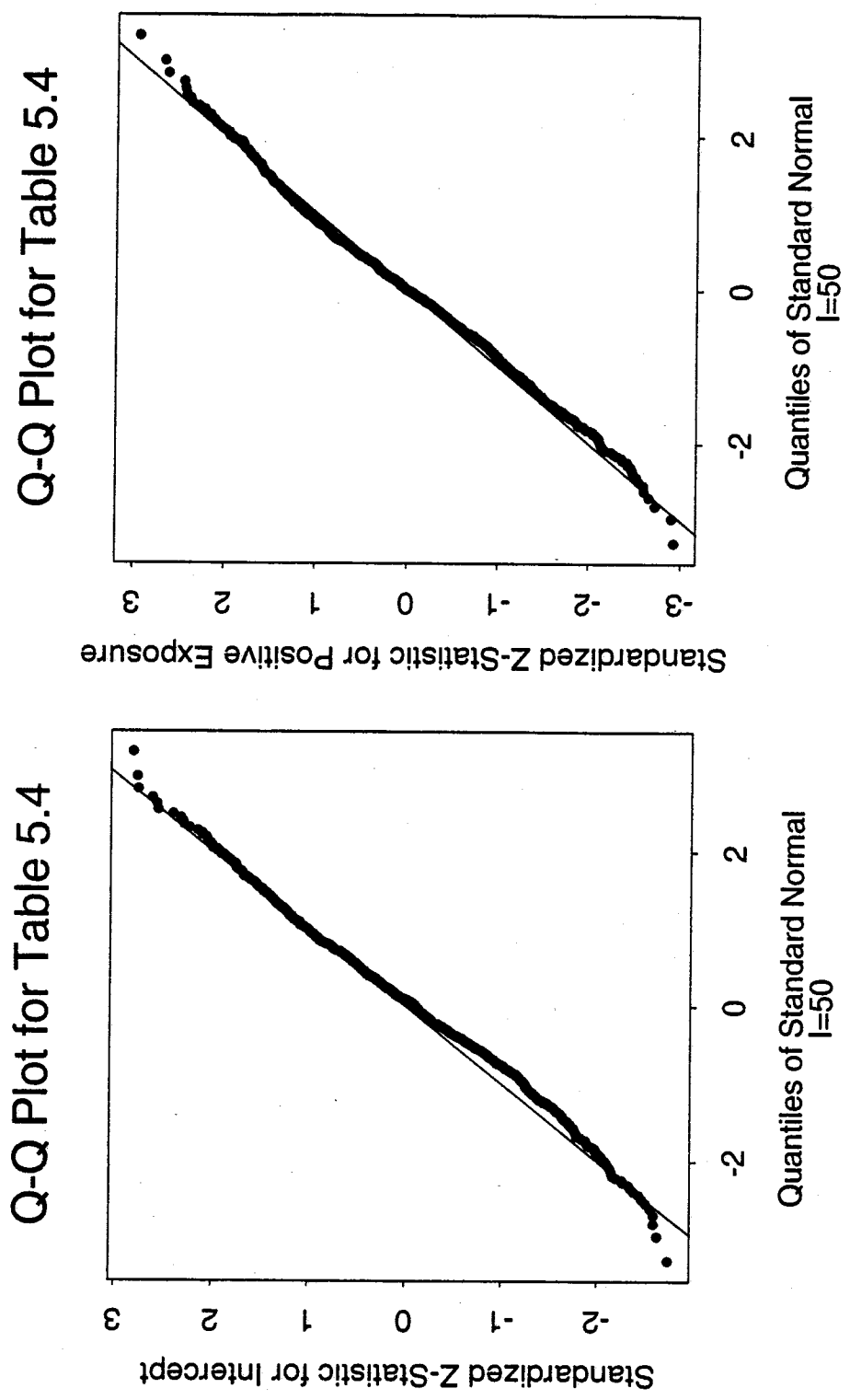

Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q=10,000, n=2
Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
Q=10,000, n=2

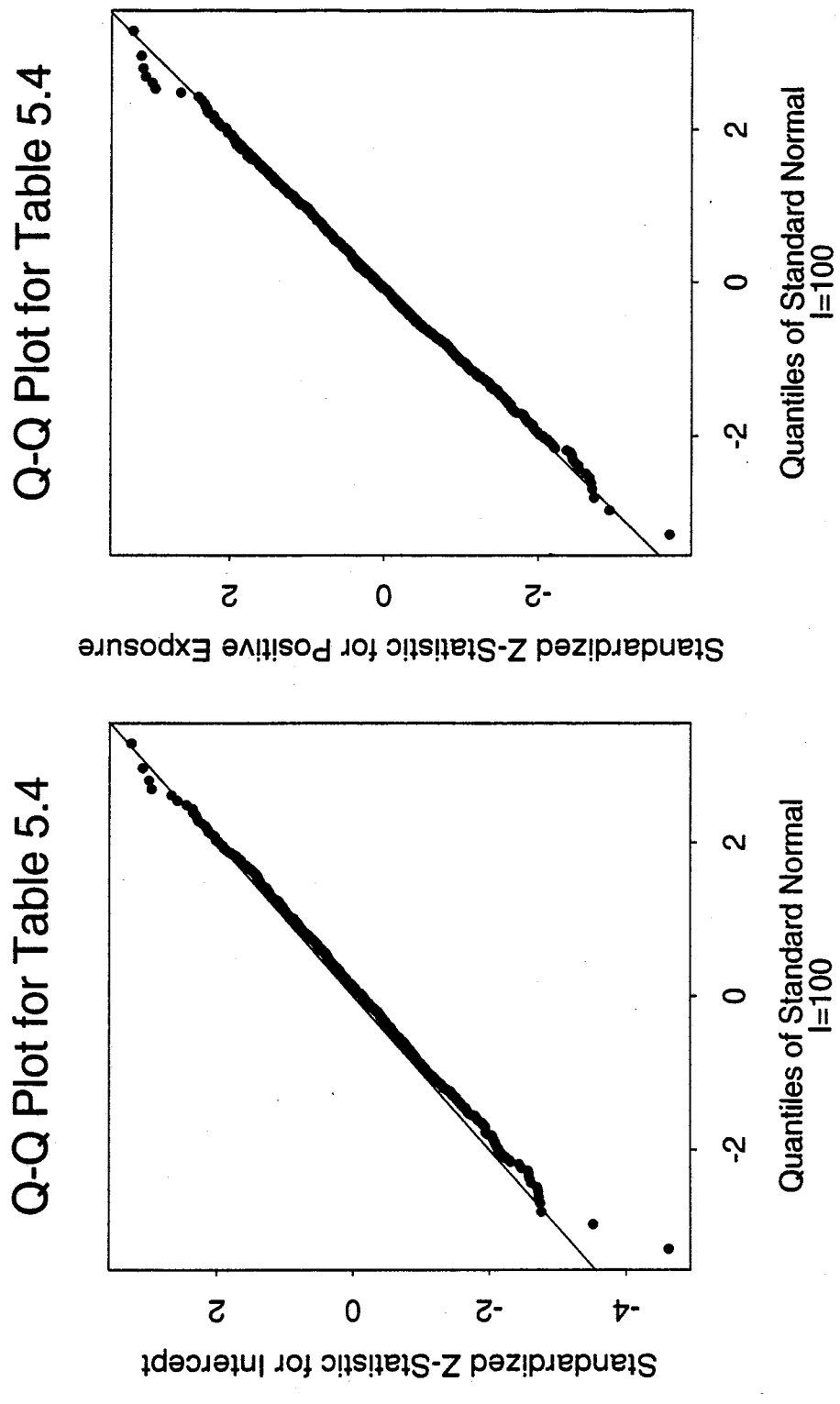

Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Intercept
3
2
1
0
-1
-2
-3
-2
0
2
Quantiles of Standard Normal
Q=50,000, n(i)<=8
Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Positive Exposure
2
0
-2
-2
0
2
Quantiles of Standard Normal
Q=50,000, n(i)<=8

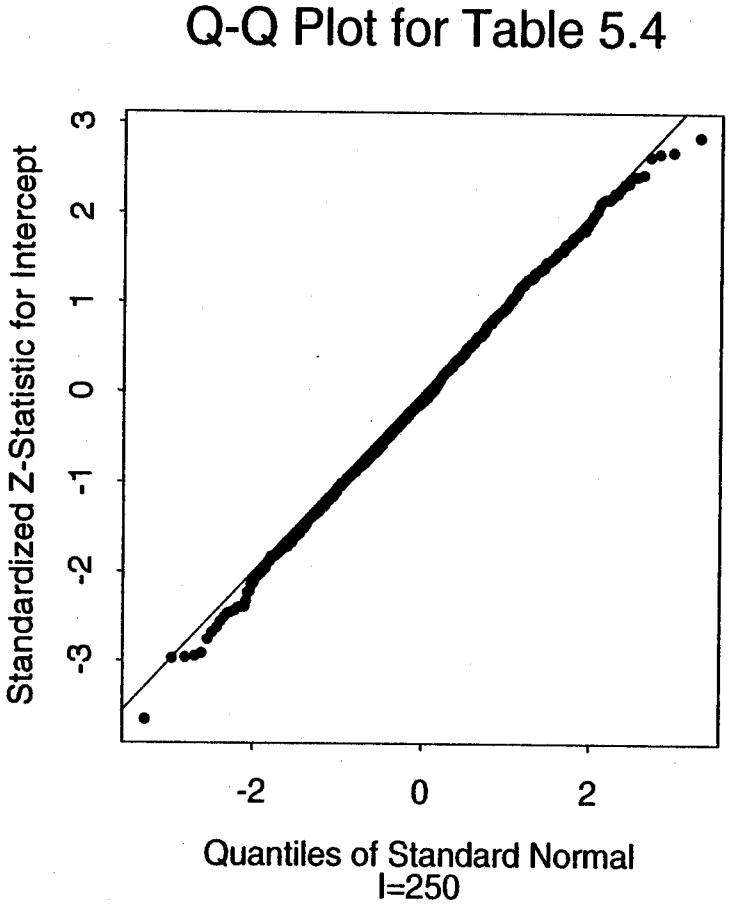

Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q=50,000, n=8
-4
-2
0
2
-2
0
2
Q-Q Plot for Table 4.21 (OS)
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
Q=50,000, n=8
-4
-2
0
2
-2
0
2

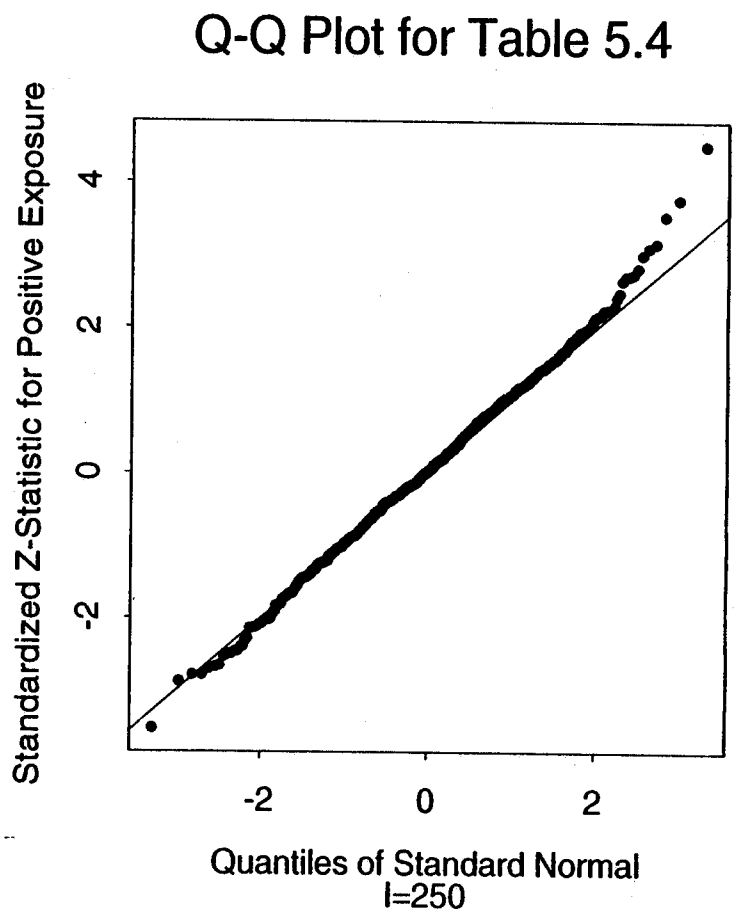

Q-Q Plot for Table 5.3
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
rho=0.2
Q-Q Plot for Table 5.3
Standardized Z-Statistic for Null Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
rho=0.2

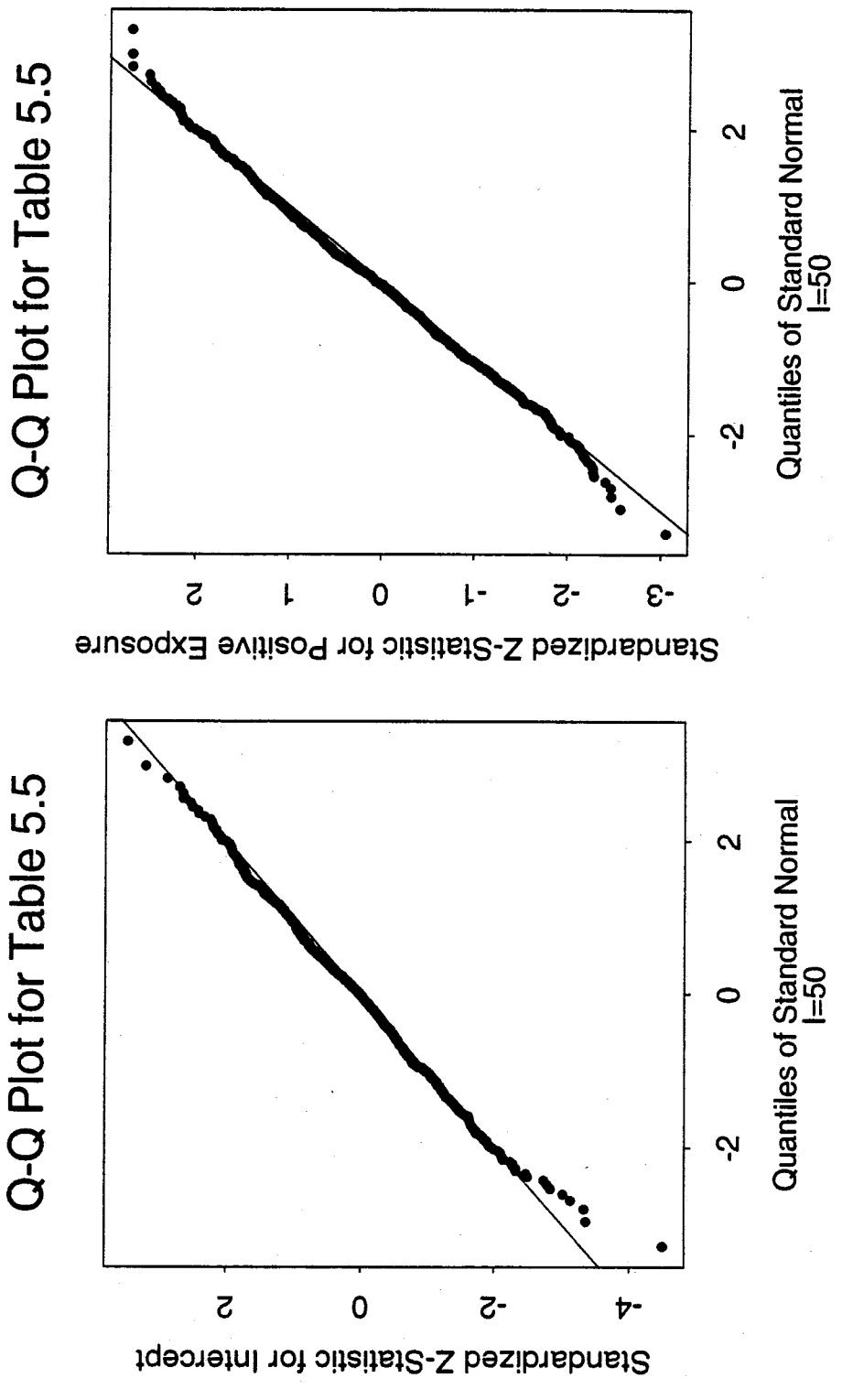

Q-Q Plot for Table 5.3
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
rho=0.4
Q-Q Plot for Table 5.3
Standardized Z-Statistic for Null Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
rho=0.4

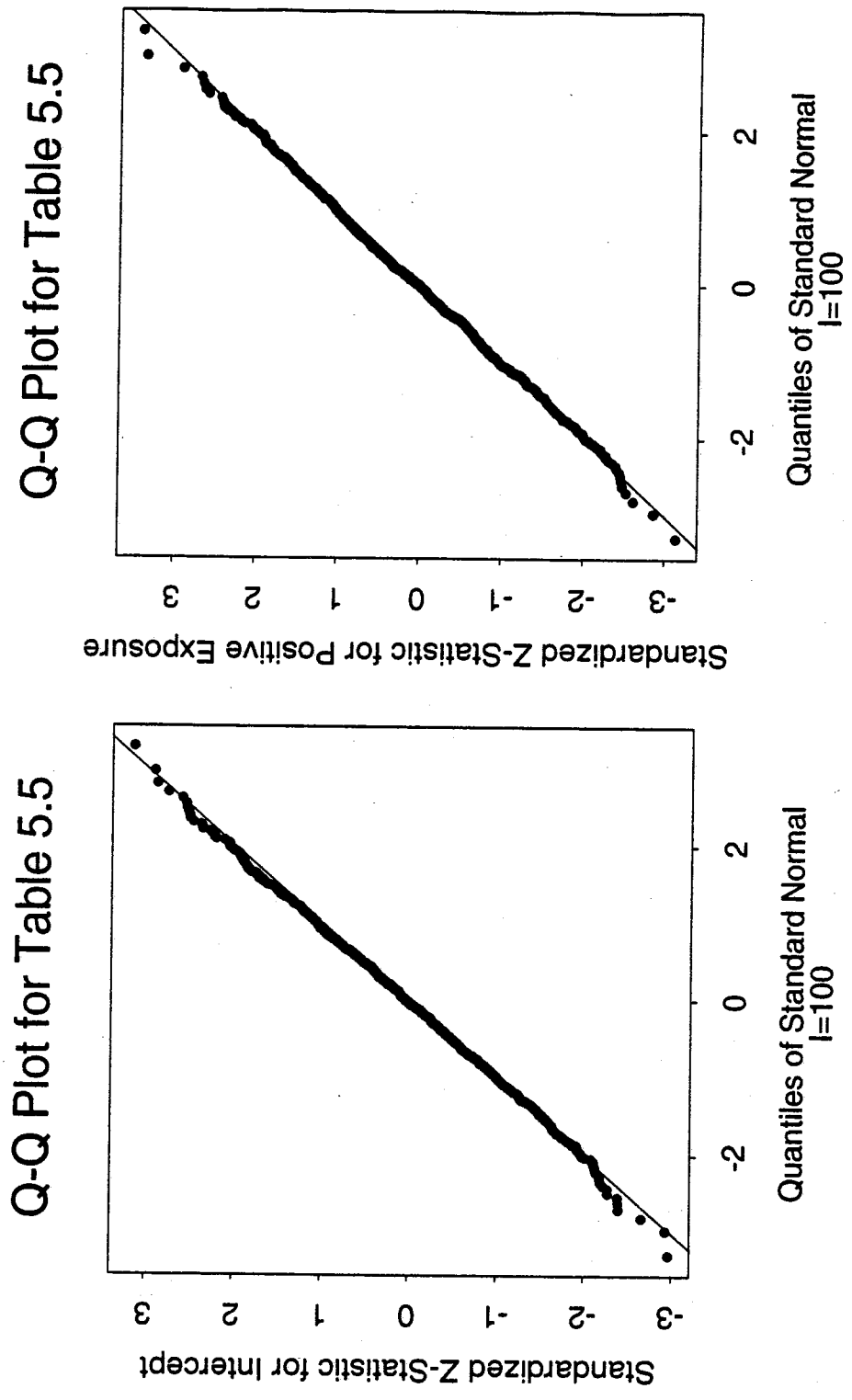

Q-Q Plot for Table 5.4
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=50
-2
-1
0
1
2
3
-2
0
2
Q-Q Plot for Table 5.4
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=50
-3
-2
-1
0
1
2
3
-2
0
2

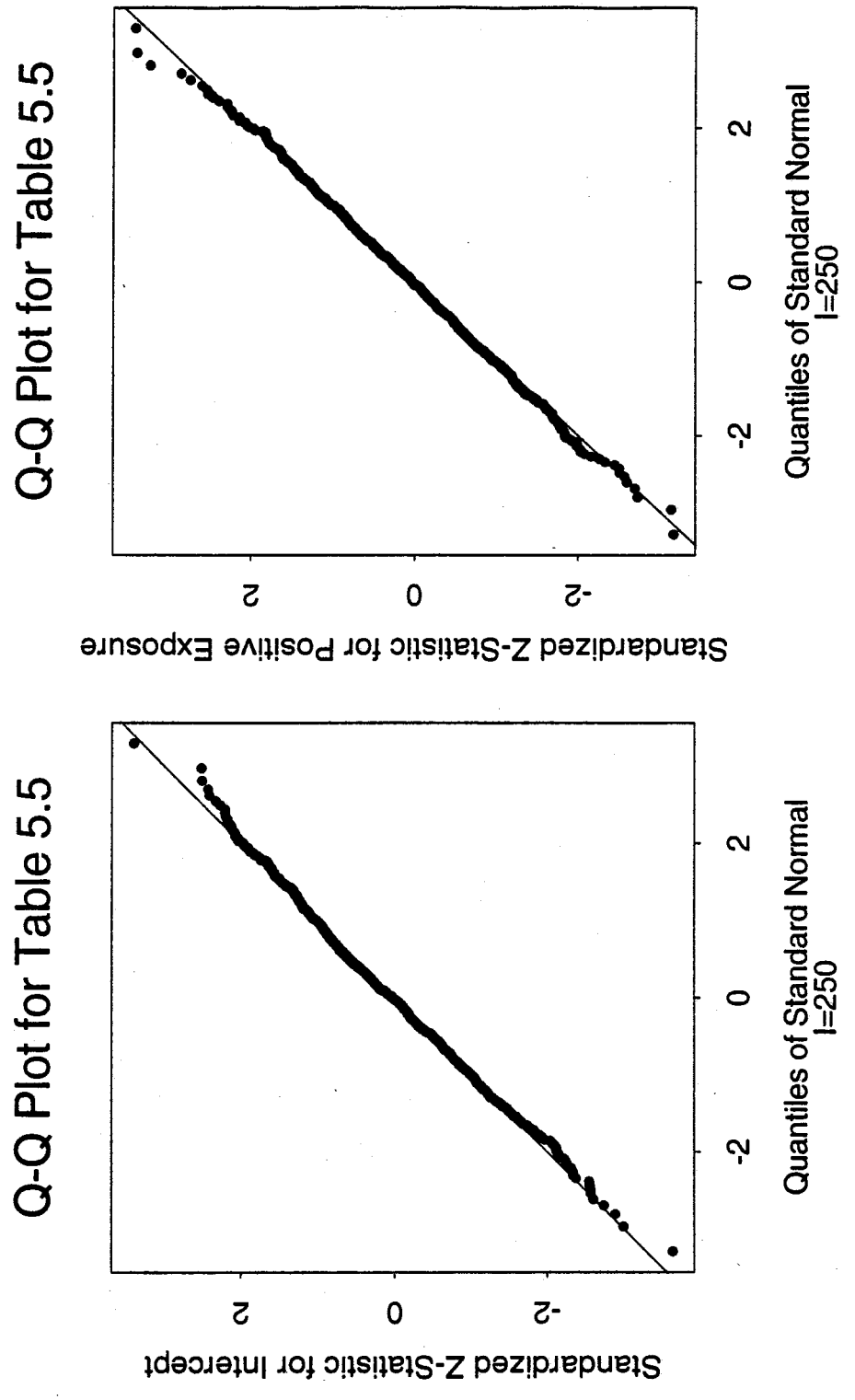

Q-Q Plot for Table 5.4
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=100
Q-Q Plot for Table 5.4
Standardized Z-Statistic for Positive Exposure
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=100

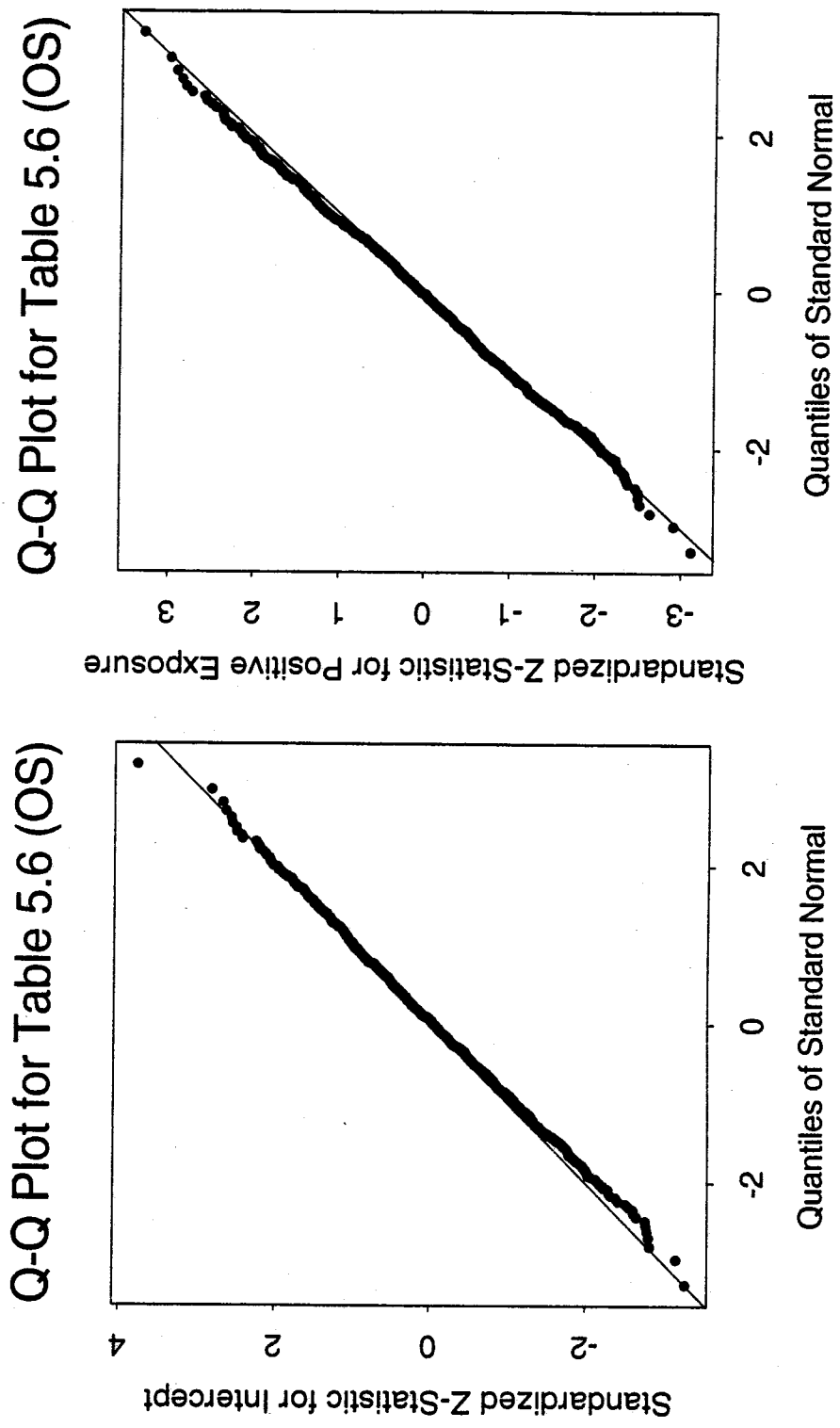

Q-Q Plot for Table 5.4
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=250

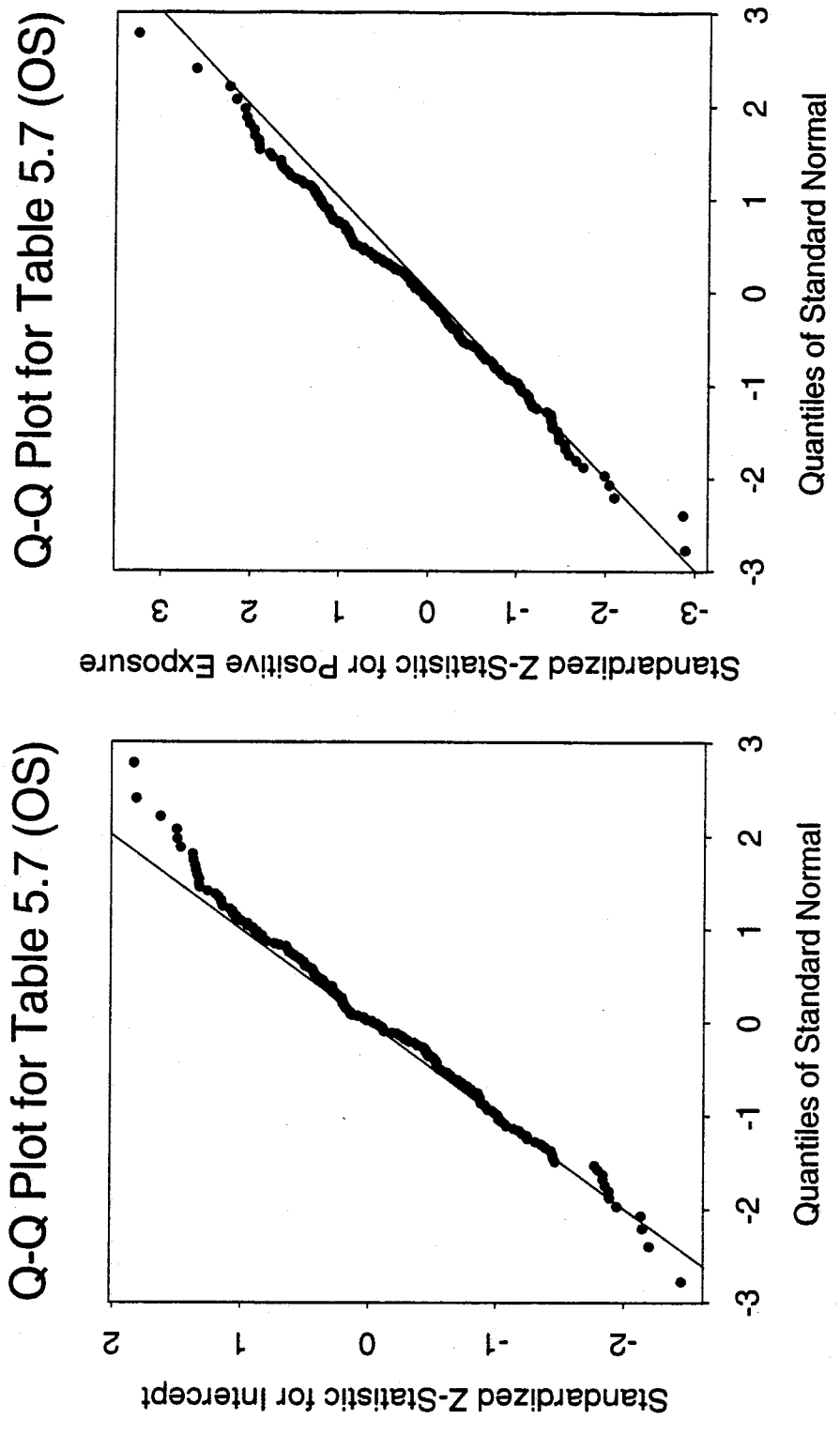

Q-Q Plot for Table 5.4
Standardized Z-Statistic for Positive Exposure
-2
0
2
4
-2
0
2
Quantiles of Standard Normal
I=250


Q-Q Plot for Table 5.5
Standardized Z-Statistic for Intercept
-4
-2
0
2
-2
0
2
Quantiles of Standard Normal
I=50
Q-Q Plot for Table 5.5
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
-2
0
2
Quantiles of Standard Normal
I=50


Q-Q Plot for Table 5.5
Standardized Z-Statistic for Intercept
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=100
Q-Q Plot for Table 5.5
Standardized Z-Statistic for Positive Exposure
-3
-2
-1
0
1
2
3
-2
0
2
Quantiles of Standard Normal
I=100


Q-Q Plot for Table 5.5
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
I=250
-2
0
2
Q-Q Plot for Table 5.5
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
I=250
-2
0
2


Q-Q Plot for Table 5.6 (OS)
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
Q-Q Plot for Table 5.6 (OS)
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal


Q-Q Plot for Table 5.7 (OS)
Standardized Z-Statistic for Intercept
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3
Q-Q Plot for Table 5.7 (OS)
Standardized Z-Statistic for Positive Exposure
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3

APPENDIX 2

MARSAGLIA RANDOM NUMBER GENERATOR PROGRAM

```
static float u[98],c,cd,cm;
static int i97,j97,set=0;
static int is1max = 31328;
static int is2max = 30081;
#include <stdio.h>

/*translation from fortran to c of marsaglia generator */
/* suggested   iseed = 54185253       */
int amrset(int iseed)
{
        int is1,is2,i,j,k,l,m,ii,jj;
        float s,t;
        is1 = iseed / is2max + 1;
        if ( is1 > is1max )is1 = is1max;
        is2 = iseed % is2max + 1;
        if ( is2 > is2max )is2 = is2max;

        i = is1 / 177;
        i = i % 177 + 2;
        j = is1 % 177 + 2;
        k = is2 / 169;
        k = k % 178 + 1;
        l = is2 % 169;
        for ( ii = 1; ii < 98; ++ii )
        {
                s = 0.0;
                t = 0.5;
                for ( jj = 1; jj < 25; ++jj )
                {
                        m = ((i*j % 179)*k) % 179;
                        i = j;
                        j = k;
                        k = m;
                        l = (53*l+1) % 169;
                        if ( (l*m % 64) >= 32 )s += t;
                        t *= 0.5;
                }

                u[ii] = s;
        }
```

```
        c  = 362436.0/16777216.0;
        cd = 7654321.0/16777216.0;
        cm = 16777213.0/16777216.0;
        i97 = 97;
        j97 = 33;
        set = 1;
}

int amrand(double *y)
{
        float uni;
        if ( !set )
        {
                fprintf(stderr,"Need to seed the rng\n");
                exit(1);
        }

        uni = u[i97] - u[j97];
        if ( uni < 0.0 )uni += 1.0;
        u[i97] = uni;
        i97 -= 1;
        if ( i97 == 0 )i97 = 97;
        j97 -= 1;
        if ( j97 == 0 )j97 = 97;
        c -= cd;
        if ( c < 0.0 )c += cm;
        uni -= c;
        if ( uni < 0.0 )uni += 1.0;
        *y = uni;
}
```

APPENDIX 3

SAMPLE C PROGRAM FOR WCR SIMULATIONS

(Observation-specific exposure, informative cluster sizes for ρ=0.2, $n_i \leq 8$)

```
#include <stdio.h>
#include <math.h>

#define iseed  99185257                    /*Seed for random number generator*/
#define logor 0.6111                       /*Cluster-specific log odds*/
/#define ncls 250                          /*Sample size*/
#define nsim 1000                          /*Number of simulations*/
#define nbeta 3
#define trueor 1.6154                      /*Marginal odds-ratio*/
#define zero 0
#define nrs 50000                          /*Number of resamples*/

main()
{
    long count=0,flag=0,flag2=0,*seed,s=-4;
    int h=0,i=0,j=0,k=0,l=0,m=0,e[ncls][8],y[ncls][8],rs[ncls],es[ncls];
    double w,rn[ncls][nbeta],min[ncls],rexp[ncls][8],unif[ncls][8],logit[ncls],
        pexp[ncls],u,beta[2],var[2];
    double beta2[2],rsvar[2],b[2],sqbeta[2],zstat[2],coverage,power,ci_low,
        ci_high,det[nsim];
    double logci_l,logci_h,b0b1,rscov,rsv[2],rsc,v[2];
    int pow,cov,notpsdef,a;
    float ytot[ncls],prob[ncls],nn[nrs][4],nn1[nrs],nn2[nrs],nn3[nrs],nn4[nrs],n[ncls];
    float ran1(long *idum);
    float bnldev(float pp, int mm, long *idum);   /*Random binomial variate generator*/
    FILE *pfile = NULL;

    amrset(iseed);
    seed=&s;
    for(h=0;h<=nsim-1;h++)                 /*Simulation loop*/
    {
        rsvar[0]=0;
        rsvar[1]=0;
        rscov=0;
        for(i=0;i<=ncls-1;i++)             /*Number of clusters*/
        {
```

```
    for(j=0;j<=nbeta-1;j++)
    {
        amrand(&w);
        rn[i][j]=w;
    }
if (rn[i][0]<=rn[i][1]) min[i]=rn[i][0];      /*Generating beta(1,3) variate*/
else min[i]=rn[i][1];

 if (min[i]<=rn[i][2]) min[i]=min[i];
else min[i]=rn[i][2];

if (min[i]>0.25)                              /*Creating informative cluster sizes*/
{
    n[i]=bnldev(0.25,9,seed);                 /*Generating unequal cluster sizes*/
}
else
{
    n[i]=bnldev(0.75,9,seed);                 /*Generating unequal cluster sizes*/
}
if (n[i]==0 || n[i]==9)                       /*Not allowing zero sized clusters*/
{
    --i;
}
else
{
    for(k=0;k<=n[i]-1 ;k++)
    {
        amrand(&w);                           /*Random number call for exposure*/
        rexp[i][k]=w;
        if (rexp[i][k]>=0.5)
        {                                     /*Observation-specific exposure*/
            e[i][k]= 0;
        }
        else
        {
            e[i][k]=1;
        }
        amrand(&w);
        unif[i][k]=w;
        if (e[i][k]==0)
        {
            if (unif[i][k]<=min[i])           /*Defining unexposed outcome values*/
            {
                y[i][k]=1;
            }
```

```
                else
                {
                    y[i][k]=0;
                }
            }
            else
            {
                logit[i]=log(min[i]/(1-min[i]))+(logor);   /*Additive term-logit scale*/
                pexp[i]=exp(logit[i])/(1+exp(logit[i]));   /*Non-null exposure effect*/
                if (unif[i][k]<=pexp[i])                   /*Outcomes for exposed*/
                {
                    y[i][k]=1;
                }
                else
                {
                    y[i][k]=0;
                }
            }                                              /*Saving data in output file*/
            pfile = fopen("filename","a");
            fprintf(pfile, "\n%4d %3d %1lf %1d
                    %1d",h+1,i+1,n[i],e[i][k],y[i][k]);
            fclose( pfile );
            }
        }
    }
    beta[0]=0;
    beta[1]=0;
    var[0]=0;
    var[1]=0;
    beta2[0]=0;
    beta2[1]=0;
    flag=0;
    sqbeta[0]=0;
    sqbeta[1]=0;
    b0b1=0;
    for(l=0;l<=nrs-1;l++)                                  /*Resampling loop*/
    {
        nn[l][0]=0;
        nn[l][1]=0;
        nn[l][2]=0;
        nn[l][3]=0;
        for(i=0;i<=ncls-1;i++)
        {
            rs[i]=0;
            es[i]=0;
```

```
amrand(&w);
u=w;
                                    /*Resampling one unit from each cluster*/
if(n[i]==1.0)                       /*Clusters of size 1*/
{
    a=1;
    rs[i]=y[i][0];
    es[i]=e[i][0];
}
if(n[i]==2.0)                       /*Clusters of size 2*/
{
    if(u < 1/n[i])
    {
        a=1;
        rs[i]=y[i][0];
        es[i]=e[i][0];
    }
    else
    {
        a=2;
        rs[i]=y[i][1];
        es[i]=e[i][1];
    }
}
if(n[i]==3.0)                       /*Clusters of size 3*/
{
    if(u < 2/n[i])
    {
        a=1;
        rs[i]=y[i][0];
        es[i]=e[i][0];
        if(u < 1/n[i])
        {
            a=2;
            rs[i]=y[i][1];
            es[i]=e[i][1];
        }
    }
    else
    {
        a=3;
        rs[i]=y[i][2];
        es[i]=e[i][2];
    }
}
```

```
if(n[i]==4.0)                    /*Clusters of size 4*/
{
    if(u < 3/n[i])
    {
        a=1;
        rs[i]=y[i][0];
        es[i]=e[i][0];
        if(u < 2/n[i])
        {
            a=2;
            rs[i]=y[i][1];
            es[i]=e[i][1];
            if(u < 1/n[i])
            {
                a=3;
                rs[i]=y[i][2];
                es[i]=e[i][2];
            }
        }
    }
    else
    {
        a=4;
        rs[i]=y[i][3];
        es[i]=e[i][3];
    }
}
if(n[i]==5.0)                    /*Clusters of size 5*/
{
    if(u < 4/n[i])
    {
        a=1;
        rs[i]=y[i][0];
        es[i]=e[i][0];
        if(u < 3/n[i])
        {
            a=2;
            rs[i]=y[i][1];
            es[i]=e[i][1];
            if(u < 2/n[i])
            {
                a=3;
                rs[i]=y[i][2];
                es[i]=e[i][2];
                if(u < 1/n[i])
```

```
                {
                    a=4;
                    rs[i]=y[i][3];
                    es[i]=e[i][3];
                }
            }
        }
    }
    else
    {
        a=5;
        rs[i]=y[i][4];
        es[i]=e[i][4];
    }
}
if(n[i]==6.0)                    /*Clusters of size 6*/
{
    if(u < 5/n[i])
    {
        a=1;
        rs[i]=y[i][0];
        es[i]=e[i][0];
        if(u < 4/n[i])
        {
            a=2;
            rs[i]=y[i][1];
            es[i]=e[i][1];
            if(u < 3/n[i])
            {
                a=3;
                rs[i]=y[i][2];
                es[i]=e[i][2];
                if(u < 2/n[i])
                {
                    a=4;
                    rs[i]=y[i][3];
                    es[i]=e[i][3];
                    if(u < 1/n[i])
                    {
                        a=5;
                        rs[i]=y[i][4];
                        es[i]=e[i][4];
                    }
                }
            }
```

```
            }
        }
        else
        {
            a=6;
            rs[i]=y[i][5];
            es[i]=e[i][5];
        }
    }
    if(n[i]==7.0)                    /*Clusters of size 7*/
    {
        if(u < 6/n[i])
        {
            a=1;
            rs[i]=y[i][0];
            es[i]=e[i][0];
            if(u < 6/n[i])
            {
                a=2;
                rs[i]=y[i][1];
                es[i]=e[i][1];
                if(u < 5/n[i])
                {
                    a=3;
                    rs[i]=y[i][2];
                    es[i]=e[i][2];
                    if(u < 4/n[i])
                    {
                        a=4;
                        rs[i]=y[i][3];
                        es[i]=e[i][3];
                        if(u < 3/n[i])
                        {
                            a=5;
                            rs[i]=y[i][4];
                            es[i]=e[i][4];
                            if(u < 2/n[i])
                            {
                                a=6;
                                rs[i]=y[i][5];
                                es[i]=e[i][5];
                            }
                        }
                    }
                }
```

```
            }
        }
        else
        {
            a=7;
            rs[i]=y[i][6];
            es[i]=e[i][6];
        }
    }
    if(n[i]==8.0)                    /*Clusters of size 8*/
    {
        if(u < 7/n[i])
        {
            a=1;
            rs[i]=y[i][0];
            es[i]=e[i][0];
            if(u < 6/n[i])
            {
                a=2;
                rs[i]=y[i][1];
                es[i]=e[i][1];
                if(u < 5/n[i])
                {
                    a=3;
                    rs[i]=y[i][2];
                    es[i]=e[i][2];
                    if(u < 4/n[i])
                    {
                        a=4;
                        rs[i]=y[i][3];
                        es[i]=e[i][3];
                        if(u < 3/n[i])
                        {
                            a=5;
                            rs[i]=y[i][4];
                            es[i]=e[i][4];
                            if(u < 2/n[i])
                            {
                                a=6;
                                rs[i]=y[i][5];
                                es[i]=e[i][5];
                                if(u < 1/n[i])
                                {
                                    a=7;
                                    rs[i]=y[i][6];
```

```
                                    es[i]=e[i][6];
                                }
                            }
                        }
                    }
                }
            }
            else
            {
                a=8;
                rs[i]=y[i][7];
                es[i]=e[i][7];
            }
        }

        if(rs[i]==1 && es[i]==1)        /*Sorting data into 2x2 table*/
        {
            ++nn[l][0];                 /*(y,e)=(1,1)*/
        }

        if(rs[i]==1 && es[i]==0)
        {
            ++nn[l][1];                 /*(y,e)=(1,0)*/
        }

        if(rs[i]==0 && es[i]==1)
        {
            ++nn[l][2];                 /*(y,e)=(0,1)*/
        }

        if(rs[i]==0 && es[i]==0)
        {
            ++nn[l][3];                 /*(y,e)=(0,0)*/
        }
}
                                        /*Check for correct sample size*/
if(nn[l][0]+nn[l][1]+nn[l][2]+nn[l][3]!=250)                {
printf("\n%lf,%lf,%lf,%lf",nn[l][0],nn[l][1],nn[l][2],nn[l][3]);
}

                                        /*finite portion of resampling distribution*/
if(nn[l][0]>0 && nn[l][1]>0 && nn[l][2]>0 && nn[l][3]>0)
```

```
    {
                                        /*sum of intercept parameters*/
        beta[0]+=(log(nn[l][1]/nn[l][3]));

                                        /*sum of exposure parameters*/
        beta[1]+=(log((nn[l][0]*nn[l][3])/(nn[l][1]*nn[l][2])));

                                        /*sum of cross products*/
        b0b1+=(log(nn[l][1]/nn[l][3])*log((nn[l][0]*nn[l][3])/(nn[l][1]*nn[l][2])));

                                        /*sum of intercepts squared*/
        beta2[0]+=(log(nn[l][1]/nn[l][3]))*(log(nn[l][1]/nn[l][3]));

                                        /*sum of exposures parms. squared*/
        beta2[1]+=(log((nn[l][0]*nn[l][3])/(nn[l][1]*nn[l][2])))*
                  (log((nn[l][0]*nn[l][3])/(nn[l][1]*nn[l][2])));

                                        /*sum of intercept variances*/
        var[0]+=((1/nn[l][1])+(1/nn[l][3]));

                                        /*sum of exposure parm. variances*/
        var[1]+=((1/nn[l][0])+(1/nn[l][1])+(1/nn[l][2])+(1/nn[l][3]));
    }
    else
    {
        ++flag;                         /*Counts number of infinite resamples*/
        ++count;
    }
}                                       /*ends resampling loop 'l' */
sqbeta[0]=beta[0]*beta[0];              /*sum of intercepts squared*/
sqbeta[1]=beta[1]*beta[1];              /*sum of exposure parameters squared*/
beta[0]=beta[0]/(nrs-flag);             /*average of finite intercept parameters*/
beta[1]=beta[1]/(nrs-flag);             /*average of finite exposure parms*/
b0b1=b0b1/(nrs-flag);                   /*average crossproduct*/
beta2[0]=beta2[0]/(nrs-flag);           /*average squared intercept*/
beta2[1]=beta2[1]/(nrs-flag);           /*average squared exposure parameter*/
var[0]=var[0]/(nrs-flag);               /*average intercept variance*/
var[1]=var[1]/(nrs-flag);               /*average exposure parameter variance*/
                                        /*intercept WCR variance*/
rsvar[0]=var[0]-(beta2[0]-(sqbeta[0]/((nrs-flag)*(nrs-flag))));
                                        /*exposure WCR variance*/
rsvar[1]=var[1]-(beta2[1]-(sqbeta[1]/((nrs-flag)*(nrs-flag))));
                                        /*WCR covariance*/
rscov=(-var[0])-(b0b1-(beta[0]*beta[1]));
if(rsvar[0]>0 && rsvar[1]>0)            /*Checking for positive WCR variances*/
```

```
{
    b[0]+=beta[0];                  /*sum of average intercepts across sims*/
    b[1]+=beta[1];                  /*sum of average exp. parms across sims*/
    rsv[0]+=rsvar[0];               /*sum of WCR intercept var. across sims*/
    rsv[1]+=rsvar[1];               /*sum of WCR exp. parm var. across sims*/
    rsc+=rscov;                     /*sum of WCR covariance across sims*/
    v[0]+=var[0];                   /*sum of average int. var. across sims*/
    v[1]+=var[1];                   /*sum of average exp. var. across sims*/
                                    /*lower limit of exp. parm 95% conf. int.*/
    ci_low=exp(beta[1]-1.96*(sqrt(rsvar[1])));
                                    /*upper limit of exp. parm 95% conf. int.*/
    ci_high=exp(beta[1]+1.96*(sqrt(rsvar[1])));

                                    /*Computing coverage*/
    if (ci_low <= trueor && trueor <= ci_high) cov=1;
    else cov=0;

    logci_l=log(ci_low);
    logci_h=log(ci_high);

                                    /*Computing power*/
    if (logci_l <= zero && zero <= logci_h) pow=0;
    else pow=1;

                                    /*Checking for pos. definite cov. matrix*/
    det[h]=(rsvar[0]*rsvar[1])-(rscov*rscov);
    if (det[h]<=0)
    {
        ++notpsdef;
        printf("\nnsim=%d,det=%lf",h,det[h]);
    }
    coverage+=cov;                  /*sum coverage across simulations*/
    power+=pow;                     /*sum power across simulations*/
    zstat[0]+=beta[0]/(sqrt(rsvar[0]));/*sum Z(int) across simulations*/
    zstat[1]+=beta[1]/(sqrt(rsvar[1]));/*sum Z(exp) across simulations*/

                                    /*Save each simulations results in file*/
    pfile = fopen("filename","a");
    fprintf(pfile, "\n%4d %2.6lf %2.6lf %2.6lf %2.6lf %2.6lf %1d
                    %1d",h+1,beta[0],beta[1],rsvar[0],rsvar[1],rscov,cov,pow);
    fclose( pfile );
}
else
{
    ++flag2;                        /*Count non-zero variance simulations*/
```

```
            det[h]=(rsvar[0]*rsvar[1])-(rscov*rscov);
            if (det[h]<=0)
            {
                ++notpsdef;
                printf("\nnsim=%d,det=%lf",h,det[h]);
            }
        }
    }
                                        /*Taking avg. of simulations with '+' var.*/
    b[0]=b[0]/((nsim)-flag2);           /*Average intercept across simulations*/
    b[1]=b[1]/(nsim -flag2);            /*Avg. exposure parameter across sims*/
    rsv[0]=rsv[0]/(nsim-flag2);         /*Average WCR int. variance across sims*/
    rsv[1]=rsv[1]/(nsim-flag2);         /*Average WCR exp. variance across sims*/
    rsc=rsc/(nsim-flag2);               /*Average WCR covariance across sims*/
    v[0]=v[0]/(nsim-flag2);             /*Average of average int. var. across sims*/
    v[1]=v[1]/(nsim-flag2);             /*Avg. of average exp. var. across sims*/
    zstat[0]=zstat[0]/(nsim -flag2);    /*Average Z(int) across simulations*/
    zstat[1]=zstat[1]/(nsim -flag2);    /*Average Z(exp) across simulations*/

                                        /*If '-' variance count as if didn't cover*/
    coverage=coverage/ nsim;            /*Average coverage across simulations*/
    power=power/nsim;                   /*Average power across simulations*/
                                        /*Save average simulation results in file*/
    printf("\nnsim=%d,rscount=%Ld,simcount=%Ld,b0=%lf,b1=%lf,rsvar0=%lf,
            v0bar=%lf,rsvar1=%lf,v1bar=%lf,rscov=%lf,z0=%lf,z1=%lf,coverage=%lf,
            power=%lf",h,count,flag2,b[0],b[1],rsv[0],v[0],rsv[1],v[1],rsc,zstat[0],
            zstat[1],coverage,power);
    printf("\n");
    pfile = fopen("filename","a");
    fprintf(pfile, "\n%4d %6Ld %6Ld %6lf %6lf %6lf %6lf %6lf %6lf
                %6lf",h,count,flag2,b[0],b[1],rsv[0],rsv[1],rsc,coverage,power);
    fclose( pfile );
}
```

APPENDIX 4

SAMPLE SAS PROGRAM FOR WCR DATA ANALYSIS

```
*****************************************;
*NAME:      ELAINE BORLAND HOFFMAN     *;
*RE:        WCR DATA ANALYSIS          *;
*          USE PROC GENMOD  IN SAS     *;
*ADVISOR:  CLARE WEINBERG              *;
*****************************************;
options nocenter ls=72 ps=58;
LIBNAME IN '~/examples';

data one;
        input lit dosage dead total;
        cards;

        /* data here */

        ;
run;

data two;
        set one;
        ratio=dead/total;
        trt=(/*define treatment here*/);
run;

proc means data=two noprint;
        var lit;
        output out=max max=count;
run;

%MACRO RESAMPLE;
%LET _NRS=_#_;                          /*Specify the number of resamples*/
%let _seed=1807;
%do _j=1 %to &_NRS;

data in.seedkeep(replace=yes);
        seed=&_seed;
        FLAG=0;
        call symput("_FLAG",PUT(FLAG,12.2));
run;
```

```
data random(replace=yes) in.seedkeep(keep=seed);
        merge in.seedkeep max;
        do m=1 to count;
                call ranuni(seed,rannum);
                CALL SYMPUT("_SEED",PUT(SEED,12.2));
                output random;
        end;
        output in.seedkeep;
run;

data three(keep=lit y trt);
        merge two random;
        if rannum le ratio then y=1;
        else y=0;
run;

proc freq data=three noprint;                   /* relevant for clustered binary data*/
                                                /* with dichotomous covariate only */
        tables y*trt / sparse missprint out=trim;
run;

proc print data=trim;
        title "Checking for zero celled data set";
run;

data trim;
        set trim;
        if count=0 then do;
                FLAG=1.00;
                call symput("_FLAG",PUT(FLAG,12.2));
        end;
run;

%IF &_FLAG=0.00 %THEN %DO;

proc genmod data=three;
        make 'ParmEst' out=out&_j;
        model y=trt / dist=? link=?;            /* choose variance and link here */
run;

data out&_j(drop=DF CHISQ PVAL);
        set out&_j;
        VAR=STDERR*STDERR;
run;
```

```
%IF &_J=1 %THEN %DO;

data out;
        SET out&_j;
run;

%END;
%ELSE %DO;

data out;
        SET out out&_j;
run;

%END;

proc datasets;
        delete trim out&_j three;
run;

%END;
%ELSE %DO;

proc datasets;
        delete trim three;
run;

%END;
%end;

proc sort data=out;
        by PARM;
run;

proc means data=out noprint;
        var ESTIMATE VAR;
        by PARM;
        output out=aver mean(estimate)=bbar mean(var)=varbar var(estimate)=se2 N=Q;
run;

data in.output_filename_here(drop=_TYPE_ _FREQ_);
        set aver;
        nrs=(&_j-1);
        rsvar=varbar-((Q-1)/Q)*se2;
run;
```

```
proc print data=in.ouput_filename_here;
run;

%MEND RESAMPLE;
%RESAMPLE;
```

REFERENCES

Altham, P.M.E. (1978). Two generalizations of the binomial distribution. *Applied Statistics*, **27**, 162-167.

Anderson, D.A. and Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B*, **47**, 203-210.

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*, London: Chapman and Hall.

Bieler, G.S. and Williams, R.L. (1995). Cluster sampling techniques in quantal response teratology and developmental toxicity studies. *Biometrics*, **51**, 764-776.

Bliss, C.I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, **22**, 134-167.

Bonney, G.E. (1987). Logistic regression for dependent binary observations. *Biometrics*, **43**, 951-973.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.

Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research, Volume 1 - The analysis of case-control studies*. Lyon, International Agency for Research on Cancer.

Carey, V., Zeger, S.L. and Diggle, P. (1993). Modeling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**(3), 517-526.

Casella, G. and Berger, R.L. (1990). *Statistical Inference*. Belmont, California: Duxbury Press.

Cessie, S.Le and Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Applied Statistics*, **43**(1), 95-108.

Chen, J.J. and Kodell, R.L. (1989). Quantitative risk assessment for teratological effects. *Journal of the American Statistical Association*, **84**, 966-971.

Chen, J.J., Kodell, R.L., Howe, R.B. and Gaylor, D.W. (1991). Analysis of trinomial responses from reproductive and developmental toxicity experiments. *Biometrics*, **47**, 1049-1058.

Conaway, M.R. (1990). A random effects model for binary data. *Biometrics*, **46**, 317-328.

Connolly, M.A. and Liang, K.Y. (1988). Conditional logistic regression models for correlated binary data. *Biometrika*, **75**, 501-506.

Cox, D.R. (1970). *The Analysis of Binary Data*. London: Methuen.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Crowder, M.J. (1978). Beta-binomial anova for proportions. *Applied Statistics*, **27**(1), 34-37.

Crowder, M.J. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82**, 407-410.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood with incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Eggenberger, F. and Polya, G. (1923). Über die Statistik verketteter Vorgänge, *Zeitschrift fur Angewandte Mathematik und Mechanik*, **1**, 179-289.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, **13**(1), 342-368.

Finney, D.J. (1971). *Probit Analysis*. (3rd edn). Cambridge University Press.

Finney, D.J. (1973). *Statistical Method in Biological Assay*. (2nd edn). New York: Hafner.

Fisher, R.A. (1935). The case of zero survivors (Appendix to Bliss, C.I. (1935)). *Annals of Applied Biology*, **22**, 164-165.

Fitzmaurice, G.M. and Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141-151.

Fitzmaurice, G.M., Laird, N.M., and Lipsitz, S.R. (1994). Analysing incomplete longitudinal binary responses: A likelihood-based approach. *Biometrika*, **50**(3), 601-612.

Galbraith, J.I. (1991). The interpretation of a regression coefficient. *Biometrics*, **47**, 1593-1595.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Gianola, D. and Foulley, J.L. (1982). Non-linear prediction of latent genetic liability with binary expression: an empirical Bayes approach. In *2nd World Congress on Genetics applied to Livestock Production*, **7**, 293-303. Madrid: Ministerio de Agricultura. Pesca y Alimentacion.

Gianola, D. and Foulley, J.L. (1983). Sire evaluation for ordered categorical data with a threshold model. *Génétique, Sélection, Évolution*, **15**, 201-224.

Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika*, **72**(3), 593-599.

Gladen, B. (1979). The use of the jackknife to estimate proportions from toxicological data in the presence of litter effects. *Journal of the American Statistical Association*, **74**, 278-283.

Gladen, B.C. (1986). On the role of 'habitual aborters' in the analysis of spontaneous abortion. *Statistics in Medicine*, **5**, 557-564.

Gnedenko, B.V. (1962). *Theory of Probability*. New York: Chelsea Publishing Company.

Harville, D.A. and Mee, R.W. (1984). A mixed-model procedure for analysing ordered categorical data. *Biometrics*, **40**, 393-408.

Haseman, J.K. and Kupper, L.L. (1979). Analysis of dichotomous response data from certain toxicological experiments. *Biometrics*, **35**, 281-294.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied logistic regression.* New York: John Wiley and Sons.

Johnson, N.L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions.* Boston: Houghton Mifflin Company.

Kupper, L.L. and Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, **34**, 69-76.

Kupper, L., Portier, C., Hogan, M. and Yamomoto, E. (1986). The impact of litter effects on dose-response modeling in teratology. *Biometrics*, **42**, 85-98.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

LaVange, L.M., Keyes, L.L., Koch, G.G. and Margolis, P.A. (1994). Application of sample survey methods for modeling ratios to incidence densities. *Statistics in Medicine*, **13**, 343-355.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1), 13-22.

Liang, K.Y., Zeger, S.L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B*, **54**(1), 3-40.

Lipsitz, S.R., Fitzmaurice, G.M., Orav, E.J. and Laird, N.M. (1994). Performance of generalized estimating equations in practical situations. *Biometrics*, **50**, 270-278.

Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1990). Maximum likelihood regression methods for paired binary data. *Statistics in Medicine*, **9**, 1517-1525.

Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, **78**, 153-160.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.

McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models, First edition*. New York: Chapman and Hall.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, Second edition*. New York: Chapman and Hall.

Moulton, L.H. and Zeger, S.L. (1989). Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics*, **45**, 381-394.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, **135**, 370-384.

Neuhaus, J.M. (1992). Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, **1**, 249-273.

Neuhaus, J.M. and Jewell, N.P. (1990). Some comments on Rosner's multiple logistic model for clustered data. *Biometrics*, **46**, 523-534.

Neuhaus, J.M., Kalbfleisch, J.D. and Hauck, W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review.* **59**(1), 25-35.

Paul, S.R. (1982). Analysis of proportions of affected foetuses in teratological experiments. *Biometrics*, **38**, 361-370.

Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. and Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review*, **64**, 89-118.

Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association*, **81**, 321-327.

Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033-1048.

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.

Qu, Y., Williams, G.W., Beck, G.J. and Goormastic, M. (1987). A generalized model of logistic regression for clustered data. *Communications in Statistics*, **16**(12), 3447-3476.

Rosner, B. (1984). Multivariate methods in ophthalmology with application to other paired-data situations. *Biometrics*, **40**, 1025-1035.

Rosner, B. and Milton, R. (1988). Significance testing for correlated binary outcome data. *Biometrics*, **44**, 505-512.

Rotnitzky, A. and Jewell, N.P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for clustered correlated data. *Biometrics*, **77**, 485-489.

Royall, R.M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review*, **54**, 221-226.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.

SAS Institute Inc. (1989). *SAS/STAT User's Guide, Version 6.* Cary, NC: SAS Institute Inc.

Seber, G.A.F. (1977). *Linear Regression Analysis.* New York: John Wiley and Sons.

Segreti, A.C. and Munson, A.E. (1981). Estimation of the median lethal dose when responses within a litter are correlated. *Biometrics*, **37**, 153-156.

Sen, P.K. and Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. New York: Chapman and Hall.

Stiratelli, R., Laird, N. and Ware, J.H. (1988). Random-effects models for serial observations with binary response. *Biometrics*, **40**, 961-971.

Stokes, M.E., Davis, C.S. and Koch, G.G. (1995). *Categorical Data Analysis Using the SAS System*, Cary, NC: SAS Institute Inc.

Stram, D.O., Wei, L.J. and Ware, J.H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association*, **83**, 631-637.

SUDAAN User's Manual, Release 6.4 (1995). Research Triangle Park, North Carolina: Research Triangle Institute.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82-86.

Ware, J.H., Lipsitz, S. and Speizer, F.E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**, 95-107.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

Wei, L.J. and Stram, D.O. (1988). Analyzing repeated measurements with possibly missing observations by modeling marginal distributions. *Statistics in Medicine*, **7**, 139-148.

Weinberg, C.R. (1993). Toward a clearer definition of confounding. *American Journal of Epidemiology*, **137**, 1-8.

Weinberg, C.R. and Gladen, B.C. (1986). The beta-geometric distribution applied to comparative fecundability studies. *Biometrics*, **42**, 547-560.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, **31**, 949-952.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, **31**, 144-148.

Williams, D.A. (1988b). Extra-binomial variation in toxicology. *Proceedings of the XIVth International Biometric Conference*, Namur, 18-23 July, 301-313.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear models: a pseudo-likelihood approach. *J. Statist. Comput. Simul.*, **48**, 233-243.

Zeger, S.L. (1988). Commentary. *Statistics in Medicine*, **7**, 161-168.

Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79-86.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121-130.

Zeger, S.L. and Liang, K.Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, **11**, 1825-1839.

Zeger, S.L., Liang, K.Y. and Albert, P. (1988). Models for longitudinal data, a generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.

Zeger, S.L., Liang, K.Y. and Self, S.G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, **72**, 31-38.

Zhao, L.P. and Prentice, R.L. (1990). Correlated binary regression using a quadratic exponential models. *Biometrika*, **77**, 642-648.