SIM 1589
pp: 1-11 (col.fig.: Nil)
PROD. TYPE: COM
ED: SNR
PAGN: VD – SCAN: Nil

STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:000–000 (DOI: 10.1002/sim.1589)

Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data

Alan Agresti*,† and Yongyi Min

Department of Statistics, University of Florida, Gainesville, FL 32611-8545, U.S.A.

SUMMARY

Binary matched-pairs data occur commonly in longitudinal studies, such as in cross-over experiments. Many analyses for comparing the matched probabilities of a particular outcome do not utilize pairs having the same outcome for each observation. An example is McNemar's test. Some methodologists find this to be counterintuitive. We review this issue in the context of subject-specific and population-averaged models for binary data, with various link functions. For standard models and inferential methods, pairs with identical outcomes may affect the estimated size of the effect and its standard error, but they have negligible, if any, effect on significance. We also discuss extension of this result to matched sets. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: binomial distribution; difference of proportions; logit model; McNemar's test; odds ratio; relative risk

1. INTRODUCTION

Matched-pairs data often occur in studies with repeated measurement of subjects, such as longitudinal studies that observe subjects over time. For a binary response, a 2×2 contingency table with the same row and column categories summarizes the data. An example is Table I, based on data from a cross-over study described by Jones and Kenward [1]. This table summarizes the results of a comparison of low-dose and high-dose analgesics for the relief of primary dysmenorrhoea.

For binary matched-pairs data, denote the number of pairs that are 'successes' for both observations by a, 'failures' for both by d, success for the first and failure for the second by b, and failure for the first and success for the second by c (see Table II). A common statistical problem with such data is to compare the probability of success for the two observations in a pair. McNemar's test of equality of the marginal success probabilities uses only b and

---

*Correspondence to: Alan Agresti, Department of Statistics, University of Florida, 204 Griffin-Floyd Hall, Box 118545, Gainsville, FL 32611-8545, U.S.A.

†E-mail: aa@stat.ufl.edu

Contract/grant sponsor: NIH
Contract/grant sponsor: NSF

Received 23 January 2003
Accepted 7 May 2003

Table I. Example of binary matched pairs from a cross-over study reported by Jones and Kenward [1].

	High dose		
Low dose	Success	Failure	Total
Success	53	8	61
Failure	16	9	25
Total	69	17	86

Table II. Notation for counts in 2×2 table for matched pairs, with proportions in parentheses.

	Column		
Row	Success	Failure	Total
Success	$a\ (p_{11})$	$b\ (p_{12})$	$a+b\ (p_1)$
Failure	$c\ (p_{21})$	$d\ (p_{22})$	$c+d$
Total	$a+c\ (p_2)$	$b+d$	n

c [2]. In χ^2 form it equals $(c-b)^2/(b+c)$. To many who first encounter this method, it seems counterintuitive that the 'concordant observations' a and d make no contribution to the test. Common intuition is that the greater the number of identical responses, the less the evidence of a true difference. For instance, in extending the test of marginal homogeneity from matched pairs to matched sets, Cochran [3] stated that 'for given values of b and c, one might feel intuitively that significance ought to be more definitely established if there are no cases in which the samples give the same result than if there are a large number of such cases'.

This note shows that, in fact, b and c provide the crucial information for such comparisons. For many analyses that seem to incorporate a and d as well, those counts may affect the estimated size of the effect and its standard error. However, they have negligible impact on its significance. Specifically, likelihood-ratio tests use only b and c, and for fixed b and c, as a and d increase the Wald statistics converge to statistics using b and c alone. The result is 'well-known' in the context of McNemar's test, but the purpose of this note is to review the issue in the broader context of models for common parameters for comparing proportions—the difference of proportions, the odds ratio and the relative risk. We also discuss extensions of this result to matched sets and to matched pairs with multi-category responses.

2. CONDITIONAL AND MARGINAL MODELS FOR MATCHED-PAIRS COMPARISONS

Denote the two observations for pair (e.g. subject) i by (y_{i1}, y_{i2}), $i = 1, \ldots, n$, where $y_{it} = 1$ is a success and $y_{it} = 0$ is a failure, $t = 1, 2$. We consider two forms of display for the data, and two related classes of models. The *subject-specific* form refers to a $2 \times 2 \times n$ table where stratum i has results for matched pair i. Row 1 contains y_{i1} in column 1 and $1 - y_{i1}$ in column 2,

Table III. Representation of four types of matched pairs.

Pair	Observation	Response	
		Success	Failure
1	First	0	1
	Second	0	1
2	First	1	0
	Second	1	0
3	First	1	0
	Second	0	1
4	First	0	1
	Second	1	0

and row 2 contains y_{i2} in column 1 and $1 - y_{i2}$ in column 2. Table III shows the four possible partial tables in this representation.

The second form is a *population-averaged* one. Averaged over the population of matched pairs, denote the probabilities of the possible response sequences (y_{i1}, y_{i2}) by π_{11} for (1, 1), π_{12} for (1, 0), π_{21} for (0, 1) and π_{22} for (0, 0). Denote the four corresponding sample proportions by $p_{11} = a/n$, $p_{12} = b/n$, $p_{21} = c/n$ and $p_{22} = d/n$. Let $\pi_1 = \pi_{11} + \pi_{12}$ and $\pi_2 = \pi_{11} + \pi_{21}$, with sample values $p_1 = (a+b)/n$ and $p_2 = (a+c)/n$. Table II summarizes the notation. The $a+d$ identical observations in Table II are the $a+d$ strata in the subject-specific table in which $y_{i1} = y_{i2}$.

We treat the cell counts in Table II as having a multinomial distribution with index n and parameters $\{\pi_{jk}\}$. The marginal totals $\sum_i y_{i1} = (a+b)$ and $\sum_i y_{i2} = (a+c)$ are dependent binomials, with index n and parameters π_1 and π_2. The row and column margins of this table are the elements in the 2×2 table that is the collapsing of the subject-specific table over the n sample pairs. We refer to Table II as the *marginal table*.

A standard model for (y_{i1}, y_{i2}) in the subject-specific table has the form

$$\text{link}[P(y_{i1} = 1)] = \alpha_i, \quad \text{link}[P(y_{i2} = 1)] = \alpha_i + \beta_\text{c} \tag{1}$$

where 'link' refers to a link function such as the logit. Model form (1) is a *conditional model*, the effect β_c being defined conditional on the pair (the c subscript in β_c refers to *conditional*). For instance, for the identity link, $P(y_{i1} = 1) = \alpha_i$ and $P(y_{i2} = 1) = \alpha_i + \beta_\text{c}$, and β_c is an assumed common difference of probabilities for each pair. Ordinary model-fitting of (1) treats (y_{i1}, y_{i2}) as independent, conditional on $\{\alpha_i\}$. With variability in $\{\alpha_i\}$, marginally (averaging over pairs) non-negative correlation occurs between the two binomial variates as summarized by np_1 and np_2 in Table II.

By contrast, let y_1 denote the first observation of a randomly selected subject and let y_2 denote the second observation of another randomly selected subject. The model

$$\text{link}[P(y_1 = 1)] = \alpha, \quad \text{link}[P(y_2 = 1)] = \alpha + \beta_\text{m} \tag{2}$$

is a *marginal model*. Its effect is population-averaged, referring to averaging over the entire population of pairs rather than to individual ones. It refers to the margins of the 2×2 population analog of Table II that cross-classifies the n values of (y_{i1}, y_{i2}), and provides comparisons of $\pi_1 = P(y_1 = 1)$ with $\pi_2 = P(y_2 = 1)$.

Before considering particular cases, we mention an argument that implies that for this marginal model, only b and c provide information relevant to testing the null hypothesis of equality of π_1 and π_2. This hypothesis is equivalent to $\beta_{\mathrm{m}} = 0$, which is itself equivalent to the symmetry structure $\pi_{21} = \pi_{12}$ in the 2×2 population marginal table. Direct calculation shows that the ratio of the maximized multinomial likelihood under $\beta_{\mathrm{m}} = 0$ in model (2) to the maximized multinomial likelihood in general depends only on b and c. The likelihood-ratio test statistic (i.e. minus two times the log of the likelihood ratio) equals $2b \log[2b/(b+c)] + 2c \log[2c/(b+c)]$ and has an asymptotic χ^2 null distribution with $\mathrm{d}f = 1$. For small samples, conditional on $b + c$, its exact null distribution is induced by a binomial distribution for c with $b+c$ trials and parameter $\frac{1}{2}$. In fact, this conditioning results naturally for the conditional model (1) with logit link, as Section 4 reviews. For Table I, this binomial test results in a two-sided exact P-value of 0.152.

The test statistic based on the large-sample normal approximation for this binomial is

$$z = \frac{c - (\frac{1}{2})(b+c)}{[(b+c)(\frac{1}{2})(\frac{1}{2})]^{1/2}} = \frac{c-b}{(b+c)^{1/2}}$$

and squaring it yields the McNemar statistic. It is the score statistic (i.e. with standard error obtained under the null) for testing marginal homogeneity. For Table I, $z = 1.63$, and the two-sided P-value equals 0.102. (This is similar to the mid P-value for the binomial test, that is, the tail probability below 8 and above 24 added to half the probability of 8 and 24, which equals 0.108. The continuity correction with the McNemar test yields $z = (15.5 - 8.5)/\sqrt{8 + 16} = 1.43$ and a P-value of 0.153, similar to the binomial with ordinary P-value. Because of operating conservatism, we will not use continuity corrections in this paper.) The Wald statistic divides the estimated effect by the non-null standard error. Its form depends on the link function for model (2), but we will see that its dependence on a and c is negligible.

3. DIFFERENCE OF PROPORTIONS

For the identity link, subject-specific and population-averaged effects are identical. For instance, for the conditional model (1), $\beta_{\mathrm{c}} = P(Y_{i2} = 1) - P(Y_{i1} = 1)$ for all i, and averaging this over pairs in the population equates β_{c} in (1) to the difference of marginal probabilities $\beta_{\mathrm{m}} = \pi_2 - \pi_1$ in model (2). Another view of this uses the result that for model (1) with this link applied to a $2 \times 2 \times n$ table, collapsibility holds when the stratum variable is marginally independent of the predictor variable [4]; here the sample cross classification of these two variables has a 1 in each cell.

The sample difference of proportions equals

$$p_2 - p_1 = p_{21} - p_{12} = (c - b)/n$$

This is the maximum likelihood (ML) estimate of β_{m} for the marginal model with identity link. For the conditional model it also results from a Mantel–Haenszel-type weighted average estimate with the subject-specific table [5]. Chen [6] showed that it is also ML for a random effects version of the conditional model in which $\{\alpha_i\}$ are randomly sampled from any parametric family. With multinomial sampling in the marginal table,

$$\mathrm{Var}(p_2 - p_1) = [(\pi_{12} + \pi_{21}) - (\pi_{21} - \pi_{12})^2]/n$$

The sample estimate of $\mathrm{Var}(p_2 - p_1)$ is

$$\widehat{\mathrm{Var}}(p_2 - p_1) = \frac{(b+c) - (c-b)^2/n}{n^2}$$

Under the null hypothesis of marginal homogeneity ($\pi_1 = \pi_2$), the variance reduces to $(\pi_{12} + \pi_{21})/n$ and its estimate to $(b+c)/n^2$, leading to the McNemar test. In the non-null case, the estimated size of the effect and its standard error diminishes as $(a+d)$ increases, since $n = (b+c) + (a+d)$. Here and in later models, we consider the effect of $(a+d)$ by letting it grow for fixed values of b and c. Then, n itself grows, and this provides a way of checking the role of a and d with asymptotics that are local to the null hypothesis. To the order n^{-2}, the estimated variance of $(p_2 - p_1)$ is then $(b+c)/n^2$, the estimated null variance. Similarly, for large n, the ratio of $(p_2 - p_1)$ to its estimated standard error is approximately $(c-b)/\sqrt{(b+c)}$, which is the standard normal form of the McNemar statistic. Thus, with the Wald test for the difference of proportions, the contribution of $(a+d)$ is minor, especially for large n. For the moderate sample size in Table I, this Wald test has a ratio of estimate to standard error of 1.66, similar to the McNemar z of 1.63.

4. ODDS RATIO

With the logit link, β in (1) and (2) refers to a log odds ratio. With the conditional model, ordinary ML estimation of β_c fails because the number of $\{\alpha_i\}$ parameters is proportional to n. The ML estimator converges in probability to double the true value [7]. A popular alternative approach eliminates $\{\alpha_i\}$ by conditioning on their sufficient statistics [8]. The resulting conditional ML estimator equals $\hat{\beta}_c = \log(c/b)$. This also equals the Mantel and Haenszel [9] estimator of a common log odds ratio for the subject-specific $2 \times 2 \times n$ table. By Neuhaus *et al.* [10], it is also the ML estimator of β_c for a random effects approach that assumes $\{\alpha_i\}$ are a random sample from a parametric distribution, as long as the pairwise correlation between the responses $(ad - bc)/\sqrt{\{(a+b)(a+c)(b+d)(c+d)\}} \geqslant 0$. The model implies a true non-negative correlation; when the sample correlation is non-negative, the observed counts in the marginal table are the marginal fit of the model, which is saturated for the marginal multinomial distribution. For the conditional ML analysis or for the random effects analysis with positive correlation, the estimated asymptotic variance of $\hat{\beta}_c$ is $(b^{-1} + c^{-1})$. Thus, with this model, a and d are irrelevant to estimation of the size of the effect or to inference about whether the effect is non-null. In particular, if this model seems plausible, this is motivation for the use of McNemar's test [8]. For Table I, $\hat{\beta}_c = \log(\frac{16}{8}) = 0.693$ has standard error 0.433, and a ratio of 1.60.

For marginal model (2) with logit link, the ML estimate of the log odds ratio is the sample log odds ratio for the margins of the marginal table,

$$\hat{\beta}_m = \log[p_2(1-p_1)/p_1(1-p_2)] = \log[(a+c)(c+d)/(b+d)(a+b)]$$

By the delta method, the estimated asymptotic variance of $\sqrt{n}(\hat{\beta}_{\mathrm{m}} - \beta)$ is

$$[p_1(1-p_1)]^{-1} + [p_2(1-p_2)]^{-1} - 2(p_{11}p_{22} - p_{12}p_{21})/[p_1(1-p_1)p_2(1-p_2)]$$

In this case, a and d influence both the estimated size of the effect and its precision. For instance, differentiation of $\hat{\beta}_{\mathrm{m}}$ with respect to a shows that $|\hat{\beta}_{\mathrm{m}}|$ decreases as a increases (or as d increases), for fixed $b \neq c$. The ratio of $\hat{\beta}_{\mathrm{m}}$ to its estimated standard error equals

$$\frac{\log(a+c) + \log(c+d) - \log(b+d) - \log(a+b)}{\sqrt{\{n[(a+d)(b+c)+4bc]/(a+b)(c+d)(a+c)(b+d)\}}}$$

For fixed $b \neq c$, suppose $a = d$. Then asymptotically as a grows, $n \approx 2a$ and the ratio i is roughly on the order of

$$a[\log(a+c) - \log(a+b)]/\sqrt{\{b+c\}} \tag{3}$$

As a increases, this converges to $(c-b)/\sqrt{\{b+c\}}$. Thus, a and d also have little influence on significance in this case.

When the conditional model holds with variability in $\{\alpha_i\}$, then $|\beta_{\mathrm{m}}| \leqslant |\beta_{\mathrm{c}}|$ [11]. Thus, $\hat{\beta}_{\mathrm{c}}$ and $\hat{\beta}_{\mathrm{m}}$ are not competitors, since the parameters differ for the two models. However, when the $\{\alpha_i\}$ in the conditional model are identical, then the marginal responses are independent and $\beta_{\mathrm{m}} = \beta_{\mathrm{c}}$. Then, the estimated asymptotic variance of $\sqrt{n}\hat{\beta}_{\mathrm{m}}$ for the marginal model simplifies to $[p_1(1-p_1)]^{-1} + [p_2(1-p_2)]^{-1}$. Straightforward calculation shows this is $\leqslant [p_1(1-p_2)]^{-1} + [p_2(1-p_1)]^{-1}$, the estimated asymptotic variance for $\sqrt{n}\hat{\beta}_{\mathrm{c}}$ under independence; see also Reference [6]. Thus, there is then some non-null efficiency loss in using the conditional estimator and ignoring the main diagonal counts. In fact, even when the two model effects are not equivalent, Liang and Zeger [12] proposed a compromise estimator of β_{c} in the conditional model that attempts to improve its efficiency by using the main diagonal counts. Their estimator smooths the usual conditional ML estimator toward the marginal model estimator, with greater weight given the marginal estimator when there appears to be less variability in $\{\alpha_i\}$. It improves efficiency while not being as biased in estimating β_{c} as the marginal estimator $\hat{\beta}_{\mathrm{m}}$ that ignores the matching. Note, however, that the efficiency loss in using $\hat{\beta}_{\mathrm{c}}$ (and hence only b and c) when the marginal responses are independent disappears as one approaches marginal homogeneity, in which case its asymptotic variance is the same as that of the marginal estimator.

5. RELATIVE RISK

With the log link function, the effect in (1) and (2) is a log relative risk. When the conditional model holds, then the marginal model has the same log relative risk, since collapsibility holds for this measure when the stratum variable is independent of the predictor variable [4]. For the marginal model, the ML estimator is the sample log relative risk for the margins of the marginal table, $\hat{\beta}_{\mathrm{m}} = \log[(a+c)/(a+b)]$. The Mantel–Haenszel-type estimate of a common log relative risk for the subject-specific table also simplifies to this [13, 14]. Chen [6] showed that it is also ML for a random effects version of the conditional model in which $\{\alpha_i\}$ are randomly sampled from any parametric family. As with the marginal odds ratio, $|\hat{\beta}_\mathrm{m}|$ decreases as a increases; however, it is unaffected by d.

By the delta method, the estimated asymptotic variance of $\hat{\beta}_\mathrm{m}$ is

$$[p_2(1-p_1)+p_1(1-p_2)-2(p_{11}p_{22}-p_{12}p_{21})]/np_1p_2=(b+c)/(a+b)(a+c)$$

Neither the estimator $\hat{\beta}_\mathrm{m}$ nor its estimated asymptotic variance depend on d. The ratio of $\hat{\beta}_\mathrm{m}$ to its estimated standard error equals

$$\frac{\log(a+c)-\log(a+b)}{\sqrt{\{(b+c)/(a+b)(a+c)\}}}$$

For fixed $b\neq c$, suppose $a=d$. Then asymptotically in a, the ratio is again on the order of (3). Thus, for large a with fixed b and c, this is approximately $(c-b)/\sqrt{\{b+c\}}$. As in other cases studied, the main diagonal counts have little influence on significance.

6. BAYESIAN INFERENCE

With logit link, the conditional model (1) and the marginal model (2) have been utilized from a Bayesian perspective. For the marginal model, Altham [15] used a Dirichlet prior for $\{\pi_{jk}\}$ with parameters $\{\mu_{jk}\}$. For the Dirichlet posterior, she noted that the posterior probability that $\pi_2>\pi_1$ equals the posterior probability that $\pi_{21}/(\pi_{12}+\pi_{21})>\frac{1}{2}$, and $\pi_{21}/(\pi_{12}+\pi_{21})$ has a beta posterior distribution with parameters $(c+\mu_{21},b+\mu_{12})$. Thus, although they affect the joint posterior distribution of (π_1,π_2), neither (a,d) nor (μ_{11},μ_{22}) affect the posterior significance. She noted that when $(\mu_{21},\mu_{12})=(1,0)$, the posterior probability that $\pi_2>\pi_1$ equals the one-sided binomial P-value for McNemar's test with the alternative that $\pi_2<\pi_1$.

By contrast, a and d do have some effect for the conditional model. Altham [15] used a prior of the form $f(\beta_\mathrm{c})\phi(\alpha_1)\cdots\phi(\alpha_n)$. When f and ϕ are symmetric about 0, she noted that the absolute difference between the posterior probability that $\beta_\mathrm{c}>0$ and $\frac{1}{2}$ is non-increasing as $a+d$ increases. However, from the form she obtained for the posterior density $f(\beta_\mathrm{c})$ of β_c, it follows that $f(\beta_\mathrm{c})/f(-\beta_\mathrm{c})=\exp[2\beta_\mathrm{c}(c-b)]$; for a given β_c, the posterior odds that $\pi_2>\pi_1$ does not depend on (a,d). The closeness to $\frac{1}{2}$ of the posterior probability that $\beta_\mathrm{c}>0$ depends strongly on the form of f. To illustrate, suppose f is discrete with probability $\frac{1}{2}$ at each of $\beta_0>0$ and $-\beta_0$. Then, applying a result from Altham [15], the posterior probability that $\beta_\mathrm{c}>0$ equals $1/\{1+\exp[2\beta_0(b-c)]\}$, which approaches $\frac{1}{2}$ as β_0 approaches 0.

Altham [15] and Ghosh *et al.* [16] considered other prior formulations for the conditional model. For instance, Ghosh *et al.* [16] used a hierarchical Bayesian approach with arbitrary link in which β_c has a flat prior, $\{\alpha_i\}$ are i.i.d. $\mathrm{N}(0,\sigma^2)$, and at the second stage σ^2 has an inverse gamma density. Then also, the absolute difference between the posterior probability that $\beta_\mathrm{c}>0$ and $\frac{1}{2}$ decreases as a or d increases.

7. EXTENSIONS TO MATCHED SETS

In summary, we have seen that with frequentist methods for standard models, pairs with identical outcomes may influence the estimated size of the effect but do not contain substantive

Table IV. Summary of results of tests for Table I and for amended table with $a = 53$ replaced by 530 and $d = 9$ replaced by 90.

		Original data			Setting $a = 530$, $d = 90$		
Parameter	Model	Estimate	SE	Ratio	Estimate	SE	Ratio
Diff. proportions	Marginal	0.0930	0.0561	1.659	0.0124	0.0076	1.636
	Conditional	0.0930	0.0561	1.659	0.0124	0.0076	1.636
Log odds ratio	Marginal	0.509	0.310	1.641	0.0932	0.0570	1.636
	Conditional	0.693	0.433	1.601	0.693	0.433	1.601
Log relative risk	Marginal	0.123	0.076	1.632	0.0148	0.0090	1.633
	Conditional	0.123	0.076	1.632	0.0148	0.0090	1.633

Note: McNemar z statistic = 1.633.

information about its significance. Table IV illustrates this. It shows results for the data in Table I for the various models studied in this note, and compares these results to what would happen if the counts of 53 and 9 for the identical outcomes were replaced by 530 and 90. Although the estimates change (except for the log odds ratio with the conditional model), so do the standard errors, and the Wald ratios approach the McNemar ratio of 1.633.

Similar results apply with matched pairs of multi-category response variables. The likelihood-ratio test of marginal homogeneity for an $I \times I$ marginal table does not depend on the I counts on the main diagonal. Likewise this is true of standard analyses for subject-specific tables, such as the multi-category extensions of the Mantel–Haenszel test [17], which are score tests, or the likelihood-ratio test comparing symmetry and quasi-symmetry models [18].

These results also extend to matched sets of $T > 2$ observations. We discuss this using the extensions of the marginal model (2) and the conditional model (1). With logit link, the marginal model for the 2^T marginal table is

$$\text{logit}[P(y_t = 1)] = \alpha + \beta_t, \quad t = 1, 2, \ldots, T \tag{4}$$

and the conditional model is

$$\text{logit}[P(y_{it} = 1)] = \alpha_i + \beta_t, \quad t = 1, 2, \ldots, T \tag{5}$$

with a constraint in each case such as $\beta_T = 0$.

For the likelihood-ratio test of marginal homogeneity (H_0: $\beta_1 = \beta_2 = \cdots = \beta_T$) for the marginal model, the value of the test statistic is completely independent of the matched sets in which all responses are successes or all responses are failures. This is also true for analogous tests for the conditional model with methods that condition on statistics that force those cell counts to be constant in the relevant sampling distribution. An example is the Cochran Q χ^2 statistic [3], which is an extension of McNemar's test from matched pairs to matched sets of observations. One can obtain that test also by conducting an extended Mantel–Haenszel test on the extension of the subject-specific Table III that has T rows for each stratum but again only one observation in each row [19]. The analysis conditions on the total success and failure totals in each stratum, and the conditional distribution is degenerate for strata in which every

Table V. Summary counts for longitudinal study of coronary risk factors in schoolchildren (from Reference [20]).

Response pattern for (1977, 1979, 1981)							
000	001	010	011	100	101	110	111
248	15	15	12	19	6	18	30

Note: 1 = obese, 0 = not obese.

Table VI. Summary of results of tests in logit model for Table V and for amended table with 300 in (0,0,0) and (1,1,1) cells.

	Original data			300 in (0,0,0), (1,1,1)		
Parameter	Estimate	SE	Ratio	Estimate	SE	Ratio
Marginal model						
$\beta_2-\beta_1$	0.034	0.122	0.28	0.012	0.042	0.28
$\beta_3-\beta_1$	−0.181	0.145	−1.29	−0.058	0.047	−1.25
$\beta_3-\beta_2$	−0.215	0.132	−1.63	−0.070	0.043	−1.64
Wald statistic	2.78			2.84		
LR statistic	2.83			2.83		
Conditional model: *Conditional ML*						
$\beta_2-\beta_1$	0.070	0.264	0.25			
$\beta_3-\beta_1$	−0.361	0.270	−1.34			
$\beta_3-\beta_2$	−0.430	0.270	−1.59			
Wald statistic	2.89					
LR statistic	2.95					
CMH statistic	2.92					
Conditional model: *Random effects ML*						
$\beta_2-\beta_1$	0.069	0.263	0.26	0.071	0.264	0.27
$\beta_3-\beta_1$	−0.365	0.272	−1.34	−0.348	0.266	−1.31
$\beta_3-\beta_2$	−0.434	0.271	−1.60	−0.419	0.266	−1.57
Wald statistic	2.90			2.83		
LR statistic	2.97			2.88		

response is a success or a failure. Likewise, conditional ML fitting of the conditional model (5) gives a likelihood that ignores matched sets of observations with identical outcomes. The conditional ML approach conditions on sufficient statistics for $\{\alpha_i\}$ in (5), which are the subject success totals, in estimating $\{\beta_t\}$. As in the matched-pairs case, the Wald statistic does depend slightly on the identical sets, but as their number increases (keeping other counts constant) the statistic converges to a limiting value that is independent of those counts.

We illustrate by applying the models to Table V, from a longitudinal study of coronary risk factors in schoolchildren [20]. A sample of children were classified by relative weight ($1=\text{obese}, 0=\text{not obese}$) in three separate years. Table VI shows the parameter estimates, standard errors, Wald statistic, and likelihood-ratio statistic for testing the time effect (H_0: $\beta_1=\beta_2=\beta_3$) for the marginal model (4) with the original data and with an amended data set that replaces the counts in cells (0,0,0) and (1,1,1) by 300 each. The likelihood-ratio statistic is constant at 2.83. The Wald statistic converges to 2.83 as the count in cells (0,0,0) and (1,1,1) increases unboundedly. (Its values for counts in cells (0,0,0) and (1,1,1) of 10^d with $d=(2,3,4,5,6)$ are (2.848, 2.833, 2.829, 2.829, 2.829).) Table VI also shows the result of the conditional ML analysis for the conditional model (5), which does not depend at all on those two cells.

Matched sets with identical outcomes can have some effect for an extension of the conditional model that provides further structure for which ML analysis uses those outcomes. Consider the random effects approach that replaces $\{\alpha_i\}$ by random intercepts $\{u_i\}$ having an $N(\alpha,\sigma^2)$ distribution. Table VI shows that the likelihood-ratio statistic for this model depends slightly on the matched sets with identical outcomes, with a slight diminution of significance as more identical triplets of observations occur. (The ML estimate of σ is 3.1 for the original data and 9.3 when the (0,0,0) and (1,1,1) cell counts are replaced by 300, and this table also illustrates how the diminution of marginal model estimates relative to conditional model estimates increases as σ increases.) As in Bayesian approaches with the conditional model, when the subject terms have a probability distribution the evidence about an effect diminishes as the subject heterogeneity increases. In addition, assuming this random effects structure for the model, the conditional ML method that ignores it (and the sets of identical outcomes) can then lose efficiency.

Therefore, for the random effects version of the conditional model, why is the matched-set case different from matched pairs in terms of the potential impact on significance tests of the identical sets? An illuminating paper in this regard is by Neuhaus and Lesperance [21]. They showed that in logit mixed-effects models with covariates, the potential loss of efficiency in ignoring the identical sets depends on the within-subject correlation. The efficiency of conditional likelihood estimators is a decreasing function of within-subject covariate correlation, and the efficiency loss can be substantial for covariates that have a strong positive within-subject correlation. However, the matched-pairs case referred to in this note has within-subject covariate correlation $=-1$, as depending on the order of viewing the observations, the dummy predictor changes from 0 to 1 or from 1 to 0. When the covariate takes its maximum negative correlation, Neuhaus and Lesperance showed that no efficiency loss occurs. For fixed covariate correlation, they also showed that the efficiency loss of conditional estimators decreases as the number of repeated observations increases, since the probability of identical observations (and, thus, discarding the subject) decreases. Interestingly, there are strong connections between conditional likelihood estimators and estimators obtained with a mixed-effects model using a non-parametric treatment of the random effects [22, 23]. Neuhaus [24] gave related efficiency comparisons of a mixed model approach with a GEE approach for a corresponding marginal model. He showed that the GEE approach using the exchangeable covariance structure suffers no efficiency loss.

We finish by mentioning a caveat, although perhaps an obvious one. Even in the matched-pairs case, one should not conclude that pairs with identical outcomes are irrelevant in *all* analyses for judging significance of effects. In more general models that add between-subject covariates, estimates of between-subject effects could be badly biased if we deleted observations with identical outcomes. For instance, suppose that when we stratify Table I by taking into account the race of the subject, all 53 (success, success) sequences are for black subjects and all nine (failure, failure) sequences are for white subjects. Then, if we deleted these 62 observations in modelling the response in terms of race and treatment, the effect of race would be drastically underestimated.

ACKNOWLEDGEMENTS

This research was partially supported by grants from NIH and NSF. The authors thank Dr Sander Greenland and a referee for helpful comments about an earlier draft.

REFERENCES

1. Jones B, Kenward, MG. Modelling binary data from a three-period cross-over trial. *Statistics in Medicine* 1987; **6**:555–564.
2. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; **12**:153–157.
3. Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950; **37**:256–266.
4. Shapiro SH. Collapsing contingency tables—a geometric approach. *American Statistician* 1982; **36**:43–46.
5. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985; **41**:55–68.
6. Chen J. On the conditional and mixture model approaches for matched pairs. *Journal of Statistical Planning and Inference* 1996; **55**:319–329.
7. Andersen EB. *Discrete Statistical Models with Social Science Applications*. North-Holland: Amsterdam, 1980; 244–245.
8. Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958; **45**:562–565.
9. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959; **22**:719–748.
10. Neuhaus JM, Kalbfleisch JD, Hauck WW. Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. *Canadian Journal of Statistics* 1994; **22**:139–148.
11. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 1991; **59**:25–35.
12. Liang KY, Zeger SL. On the use of concordant pairs in matched case-control studies. *Biometrics* 1988; **44**: 1145–1156.
13. Nurminen M. Asymptotic efficiency of general noniterative estimators of common relative risk. *Biometrika* 1981; **68**:525–530.
14. Tarone RE. On summary estimators of relative risk. *Journal of Chronic Diseases* 1981; **34**:463–468.
15. Altham PME. The analysis of matched proportions. *Biometrika* 1971; **58**:561–576.
16. Ghosh M, Chen M, Ghosh A, Agresti A. Hierarchical Bayesian analysis of binary matched pairs data. *Statistica Sinica* 2000; **10**:647–657.
17. Landis JR, Heyman ER, Koch GG. Average partial association in three-way contingency tables: A review and discussion of alternative tests. *International Statistical Review* 1978; **46**:237–254.
18. Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis*. MIT Press: Cambridge, MA, 1975; 307.
19. White AA, Landis JR, Cooper MM. A note on the equivalence of several marginal homogeneity test criteria for categorical data. *International Statistical Review* 1982; **50**:27–34.
20. Woolson RF, Clarke WR. Analysis of categorical incomplete longitudinal data. *Journal of the Royal Statistical Society*, *Series A* 1984; **19**:251–253.
21. Neuhaus JM, Lesperance ML. Estimation efficiency in a binary mixed-effects model setting. *Biometrika* 1996; **83**:441–446.
22. Lindsey B, Clogg CC, Grego G. Semi-parametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* 1991; **86**:96–107.
23. Tjur T. A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics* 1982; **9**:23–30.
24. Neuhaus JM. Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics* 1993; **49**:989–996.