REVSTAT – Statistical Journal
Volume 18, Number 5, October 2020, 581–592

MATCHED PAIRS WITH BINARY OUTCOMES

Authors: Christiana Kartsonaki
– MRC Population Health Research Unit, Nuffield Department of Population Health,
University of Oxford,
Oxford OX3 7LF, UK
christiana.kartsonaki@dph.ox.ac.uk

D.R. Cox
– Nuffield College,
Oxford OX1 1NF, UK
david.cox@nuffield.ox.ac.uk

Received: September 2018 Revised: May 2019 Accepted: June 2019

Abstract:

- Comparison of two treatments in matched pairs is a powerful general method for improving precision. When the outcome is binary the formulation in terms of logistic comparisons leads to an analysis in which concordant pairs, that is pairs in which both members show the same outcome, are discarded. The present paper discusses a number of conceptual aspects of this including a comparison with a linear in probabilities formulation, the relation between logistic parameters in different designs and in particular some new efficiency comparisons. Some emphasis is based on new relations between estimated effects derived from different formulations and on comparative calculations of asymptotic efficiency.

Key-Words:

- *asymptotic relative efficiency; comparison of distinct models; log odds; logistic regression.*

AMS Subject Classification:

- 62BO5 : 62F∗∗.

1. INTRODUCTION

The comparison of two treatments, T_0 and T_1, using matched pairs of individuals is a simple and often effective way of improving precision and is the basis for many generalizations. When each outcome is binary, 0 or 1, say, there are four possible outcomes from a pair, (0, 0) and (1, 1), called concordant pairs, and (1, 0) and (0, 1), called discordant pairs. The analysis of such data has been extensively discussed, partly because of the broader implications for approaching formalized statistical inference; see, for example, the wide-ranging review of Agresti [1] (1990).

McNemar [10] (1947) suggested testing the null hypothesis of treatment equivalence by discarding the concordant pairs and testing the discordant pairs for an equal split between the two possibilities, using the binomial distribution with parameter 1/2. Cox [3, 4] (1958a, b) formalized this within a systematic approach to the analysis of binary data using an exponential family setting based on a linear logistic model. In the psychometric literature the problem is considered in this way as a special case of the Rasch model (Rasch [12], 1960).

One approach, possibly closer in spirit to McNemar's paper, is to treat the analysis as a simple significance test (Fisher [8, Chapter II], 1935) in which the strong null hypothesis is that the outcome on each individual is totally unaffected by the treatment allocations, taken to be by design independent randomization for each pair between the two possible assignments. With m pairs and two treatments there are thus 2^m possible configurations that might be observed, each with the same probability under the null hypothesis; McNemar's test follows from that. Here a stochastic model for the data is not needed; rather the stochastic element comes from the randomization in design. This is a powerful argument but limited in its implications because an estimation formulation attached to it is rather contrived. The extensive literature on the matched pairs and related issues has tended not to follow that route.

A general aspect that underlies the discussion of binary data goes back in particular to earlier differences between Karl Pearson and Yule (Pearson [11], 1907; Yule [13], 1903). The former treated the two binary variables in a simple 2 × 2 contingency table as derived from an underlying standardized bivariate Gaussian distribution whose correlation coefficient is to be estimated, whereas Yule considered the binary variables as such.

In the logistic formulation, let Y_{s0}, Y_{s1} be independent random variables representing the observations on the sth pair and suppose that for m pairs

$$P(Y_{s0} = i) = L_i(\alpha_s - \theta/2), \quad P(Y_{s1} = j) = L_j(\alpha_s + \theta/2), \qquad i, j \in \{0, 1\}, \tag{1.1}$$

where $\alpha_s, \theta \in \mathbb{R}$ are unknown parameters, $L_1(x) = e^x/(1 + e^x)$ is the unit logistic function and $s = 1, ..., m$ and we write $L_0(x) = 1 - L_1(x)$. Interest is typically focused on θ whereas the α_s specify inter-pair differences supposed to be of no direct interest. Here i, j take values 0 and 1 and the parameter space is unconstrained.

It follows from the existence of complete sufficient statistics that if study of θ is to be made regarding the α_s as totally arbitrary nuisance parameters, then to achieve a procedure not formally depending on those parameters, inference is made conditionally on the pair totals, and therefore confined to the discordant pairs. In the nonnull case this leads to confidence limits for θ based on the binomial distribution.

The discarding of concordant pairs has often been regarded with unease, especially if there are many such pairs, this superficially pointing to treatment equivalence. If, though, it is required to cover the possibility that many of the α_s are large in absolute value the rejection seems inevitable. If, however, implicit or explicit restrictions are placed on the variation of these parameters some information may be recoverable from the concordant pairs. For example, Lee [9] (2001) replaced the logistic model by a broadly equivalent Poisson model. Because of the richer reference set somewhat improved estimates involving the concordant pairs were obtained.

Another route is to replace the logistic function in the above formulation by some other function, for example the linear or Gaussian functions. Such a change might give a better fit or a more direct interpretation or accommodate several related studies more conformably. Aranda-Ordaz [2] (1981) studied a parametric family of transformations as a basis for choosing the best fitting model. Empirical discrimination betweeen different models typically requires extensive data.

In essentially discrete problems "exact" significance testing involves reference to a discrete distribution and hence for each data configuration to a limited set of achievable significance levels. There is a very extensive literature on how the discrete test can be augmented to achieve some pre-specified level, such as 0.05. For interpretative purposes such arbitrarily defined levels are irrelevant. Repetitive binary decision problems such as routine screening need to be treated as such.

2. OUTLINE ANALYSES OF TWO MODELS

For data from m independent pairs we write for pair s the likelihood contribution for outcome (i, j) as

$$L_i(\alpha_s - \theta/2)L_j(\alpha_s + \theta/2).$$

It follows that, if $\hat{\pi}_{ij}$ is the proportion of pairs with $Y_0 = i, Y_1 = j$ with corresponding probabilities π_{ij}, then

$$\pi_{ij} = \text{Ave}_s L_i(\alpha_s - \theta/2)L_j(\alpha_s + \theta/2), \tag{2.1}$$

where s indexes the pairs and Ave_s is the average over the m pairs in the study or over a population of pairs from which the observed pairs have been randomly chosen. For each fixed θ the sufficient statistics for the α_s are the pair totals. Conditioning on these leaves no information in the concordant pairs and the contribution to the conditional log likelihood is thus $L_1(\theta)$ from each of the N_{01} pairs with outcome (0, 1) and $L_0(\theta)$ from each of N_{10} pairs (1, 0). It follows that $\hat{\theta} = \log(N_{01}/N_{10})$ and $L_1(\hat{\theta}) = N_{01}/(N_{01} + N_{10})$, so that, in particular, from the variance of a binomial distribution,

$$\text{var}\{L_1(\hat{\theta})\} = L_0(\theta)L_1(\theta)/(m\pi_D), \tag{2.2}$$

and then by the formula for the asymptotic variance of nonlinear function

$$\mathrm{var}(\hat{\theta}) = 1/\{mL_0(\theta)L_1(\theta)\pi_D\},$$

where π_D is the average probability that a pair is discordant so that $m\pi_D$ is the expected number of discordant pairs.

Suppose now that we replace the logistic formulation by the linear representation

$$P(Y_{s0}=1) = 1/2 + \beta_s - \phi/2,\ P(Y_{s1}=1) = 1/2 + \beta_s + \phi/2, \tag{2.3}$$

where β_s specifies the impact of the sth pair and ϕ gives the difference of probabilities between the two groups. The parameter space is constrained so that all probabilities are in [0, 1]. This places relatively complicated restrictions on the component parameters. If we write $\mu_\beta = \Sigma\beta_s/m,\ \ \sigma_\beta^2 = \Sigma\beta_s^2/m - \mu_\beta^2$, then the four cell probabilities for the expected outcome proportions are in a symmetrized notation

$$\begin{aligned}
\pi_{00} &= 1/4 - \phi^2/4 - \mu_\beta + \gamma_\beta,\\
\pi_{01} &= 1/4 + \phi/2 + \phi^2/4 - \gamma_\beta,\\
\pi_{10} &= 1/4 - \phi/2 + \phi^2/4 - \gamma_\beta,\\
\pi_{11} &= 1/4 - \phi^2/4 + \mu_\beta + \gamma_\beta,
\end{aligned}$$

where $\gamma_\beta = \mu_\beta^2 + \sigma_\beta^2$. Here ϕ specifies the inter-treatment differences and β_s characterizes the sth pair. Explicit characterization of the parameter space, that is the non-negativity of p_{ij} is not simple.

It follows that $\phi = \pi_{01} - \pi_{10}$ is estimated by

$$\hat{\phi} = \hat{\pi}_{01} - \hat{\pi}_{10} = (N_{01} - N_{10})/m. \tag{2.4}$$

The numerator is the sum of independent random variables taking the values $(-1, 0, 1)$ and it follows that

$$\mathrm{var}(\hat{\phi}) = \frac{1 - \phi^2 - 4\gamma_\beta}{2m} = (\pi_D - \phi^2)/m. \tag{2.5}$$

This depends not only on the discordant pairs but, through the denominator, also on the total number of concordant pairs.

The variance component σ_β^2 can be estimated through its equivalence to

$$1/4 - (\pi_{0.} - \pi_{1.})(\pi_{.0} - \pi_{.1})/4 - \pi_D/2.$$

In this discussion μ_β and σ_β^2 are the mean and variance of the finite population of values of β_s. Alternatively if the β_s correspond to independent and identically distributed random variables and expectations are taken over their distribution the parameters μ_β and σ_β^2 refer to that distribution.

3. SOME SIMPLE COMPARISONS

Comparison of logistic and linear in probability and indeed other models can be viewed in a number of distinct ways. From the viewpoint of formal statistical theory the logistic model has the major advantage of leading to a full exponential family form with the regression coefficients as canonical parameters (Cox [3], 1958a) and associated "exact" methods. Fully efficient estimation for the linear in probability models requires iterative calculation. However the use of ordinary least squares, treating the binary (0, 1) outcomes as if quantitative, has high efficiency so long as the probabilities are in a central range, say $(0.2, 0.8)$ (Cox and Wermuth [7], 1992). The direct subject-matter interpretation of differences in probabilities in terms of expected numbers of individuals affected is an advantage of the linear in probabilities model but the severe restrictions to specified regions of the parameter space are a major disadvantage of that formulation.

There is, however, a further general consideration applying to all issues connected with binary data and going back to the early work of Karl Pearson [11] (1907) and Yule [13] (1903) on the simpler 2×2 table. Pearson treated binary variables as formed from dichotomizing unobserved continuous variables having a bivariate normal distribution whose correlation is the focus of interest, whereas Yule treated binary variables directly in their own right. In many contexts the distinction is nugatory, although for quantal bioassay the former approach is directly relevant. Each study individual has a just critical dose above which, say, a lethal response is observed; each individual can be tested only once. Treating the unobserved critical dose levels as having a normal distribution, virtually indistinguishable from a continuous logistic distribution, is often reasonable; treating it as uniform, the implication of the linear in probabilities model, would typically not be.

The distinction between logistic and linear formulation disappears at the null hypothesis $\theta = \phi = 0$ and locally the parameter estimated in the linear in probability model is $\phi = \text{Ave}_s\{L_1(\alpha_s + \theta/2) - L_1(\alpha_s - \theta/2)\} = \theta \text{Ave}_s\{L_1'(\alpha_s)\}$, where $L_1'(\cdot)$ is the derivate of $L_1(\cdot)$, and this is approximately

$$\tilde{\phi} = \theta \text{Ave}_s\{L_1(\alpha_s)L_0(\alpha_s)\} = \theta\pi_D/2.$$

Here $\text{Ave}_s(b_s)$ is the unweighted average $\Sigma b_s/m$. The asymptotic relative efficiency of the linear and logistic procedures is thus given by the ratio $\text{var}(\tilde{\phi})/\text{var}(\hat{\phi})$ evaluated at the null hypothesis and this is one.

Both logistic and linear formulations have three free parameters and are therefore saturated families for the distribution over the four possible outcomes. The linear in probability model has for most purposes the more directly understandable interpretation, although if the proportions of, say, 1's are small, the interpretation of the logistic model in terms of proportional effects is attractive and the positivity constraints on the linear model are severe. Often the most appealing base for choosing between different formulations is stability of estimated effects across replicate sets of data, that is relative constancy of either θ or of ϕ, potentially favouring the logistic formulation.

Instead of matching in pairs it would be possible to randomize the allocation of individuals to the two groups, leading to a comparison of two binomially distributed random variables. We study the consequences of this in Section 8.

4. SOME APPROXIMATIONS

A number of aspects of the study of logistic models involve the evaluation of expectations typified in its simplest form by $E\{L_1(\mu + A)\}$, where A is a random variable of zero mean and variance σ^2. There are a number of approximations for small σ equivalent to order σ^2 but one that is likely to be better over a wider range of values. The simplest is based on Taylor expansion of $L_1(\mu + A)$ for small A and is

$$E\{L_1(\mu + A)\} = L_1(\mu) + \sigma^2 L_1''(\mu)/2, \tag{4.1}$$

where

$$L_1''(\mu) = L_1(\mu)L_0(\mu)\left\{L_0(\mu) - L_1(\mu)\right\}.$$

The second approximation is based on absorbing the correction term in (4.1) into the first by writing the approximation

$$L_1\left\{\mu + \sigma^2\{L_0(\mu) - L_1(\mu)\}/2\right\},$$

differing from (4.1) by terms of order $O(\sigma^4)$.

A third approximation is obtained less directly but is more stable for larger values of σ^2. We approximate the logistic function $L_1(x)$ by the standardized normal integral $\Phi(kx)$ for a suitable constant k; this gives a good approximation over a wide range of arguments. Then the expectation of interest is approximately $E\{\Phi(k\mu + kA)\}$ and if also A is normally distributed this expectation is itself a normal integral. On re-expressing this as a logistic function we obtain the third approximation

$$L_1\left\{\frac{\mu}{(1 + k^2\sigma^2)^{1/2}}\right\}.$$

Suitable values of k are suggested by Cox and Snell [6, p. 21–22] (1989); a compromise value over the central part of the range is $k = 0.607$. A major advantage of this third approximation is that, unlike the other two, it gives qualitatively sensible answers even for large values of σ^2.

To aid interpretation, suppose for instance that the probabilities varied with roughly 95% of values being between 0.6 and 0.9. Then the corresponding logistic function varies between 0.4 and 2.2 suggesting a σ of roughly 0.45. Then the correction factor $\sqrt{\{1 + (0.607^2 \cdot 0.45^2)\}}$ would be about 1.04, implying a quite modest adjustment.

For more detailed comparisons more explicit information about the probability that a pair is discordant is needed. We treat α_s as a random variable A, so that

$$\pi_D = E_A\left\{L_0(\mu + A - \theta/2)L_1(\mu + A + \theta/2) + L_1(\mu + A - \theta/2)L_0(\mu + A + \theta/2)\right\}. \tag{4.2}$$

This can be expanded in terms of σ by the methods outlined above. The complex details will not be given.

Table 1 shows π_D against θ, μ and σ. The calculated values of π_D were confirmed by simulation. The proportion of discordant pairs decreases rather slowly with σ.

Table 1: Probability of a pair being discordant, π_D, against θ, μ and σ.

		$\pi_D(0)$				
	θ	0	0.5	1	1.5	2
σ	μ					
	0	0.500	0.530	0.607	0.702	0.790
	0.5	0.470	0.500	0.578	0.676	0.769
	1	0.393	0.422	0.500	0.604	0.709
	1.5	0.298	0.324	0.396	0.500	0.615
		$\pi_D(\sigma)$ / $\pi_D(0)$				
0.5	0	0.94	0.94	0.95	0.96	0.97
	0.5	0.95	0.95	0.96	0.97	0.97
	1	0.98	0.98	0.97	0.97	0.98
	1.5	1.01	1.01	1.00	0.99	0.98
	2	1.05	1.04	1.03	1.01	1.00
1	0	0.75	0.76	0.80	0.85	0.90
	0.5	0.79	0.80	0.83	0.86	0.90
	1	0.91	0.90	0.90	0.89	0.90
	1.5	1.05	1.04	1.00	0.95	0.93
	2	1.19	1.17	1.11	1.04	0.98
1.5	0	0.44	0.47	0.56	0.66	0.76
	0.5	0.54	0.56	0.61	0.69	0.77
	1	0.80	0.79	0.76	0.76	0.78
	1.5	1.12	1.08	0.99	0.90	0.84
	2	1.42	1.37	1.25	1.10	0.96

Estimation of σ is in principle possible by first estimating θ and μ and then comparing the proportion of discordant observations with that to be expected in the homogenous case, $\sigma = 0$. Table 1 shows that it is only for rather large value of σ and even then for certain ranges of the other parameters that such estimation is likely to be effective.

5. UNCONDITIONAL ANALYSIS

Suppose that instead of pairing, individuals are randomized to two groups, 0 and 1, therefore with probabilities of success

$$P(Y_0 = 1) = E\left\{L_1(\mu + A - \theta/2)\right\}, \quad P(Y_1 = 1) = E\left\{L_1(\mu + A + \theta/2)\right\},$$

respectively. The resulting unconditional analysis uses all pairs.

Thus, for example, the probability of success for an individual in group 0 is approximately

$$\psi_0 \simeq L_1\left(\frac{\mu - \theta/2}{\surd(1 + k^2\sigma^2)}\right) \tag{5.1}$$

and that for an individual in group 1 is

$$\psi_1 \simeq L_1\left(\frac{\mu + \theta/2}{\surd(1 + k^2\sigma^2)}\right). \tag{5.2}$$

To estimate the marginal log odds ratio, we calculate

$$\operatorname{logit}(\psi_1) - \operatorname{logit}(\psi_0) = \frac{\theta}{\surd(1 + k^2\sigma^2)}.$$

Thus the sample proportions can be used to obtain an unconditional estimate of θ

$$\hat{\theta}_U = \surd(1 + k^2\sigma^2)\left\{\log\frac{\hat{\psi}_1}{1-\hat{\psi}_1} - \log\frac{\hat{\psi}_0}{1-\hat{\psi}_0}\right\},$$

where in this discussion we shall treat σ as known or, more realistically, treated by sensitivity analysis.

The asymptotic variance of the estimate of the treatment effect θ in the unconditional analysis is then

$$\operatorname{var}(\hat{\theta}_U) \simeq (1 + k^2\sigma^2)\left\{\frac{1}{m\psi_0(1-\psi_0)} + \frac{1}{n\psi_1(1-\psi_1)}\right\}, \tag{5.3}$$

which can be expressed in terms of the functions L_i. The parameter σ^2 might possibly be estimated from the proportion of discordant pairs, although the resulting precision is likely to be low.

Table 2 shows $\operatorname{var}(\hat{\theta}_U)$ against θ, μ and σ. The variance of the estimate of θ from the unconditional analysis increases with μ. The relation between σ and the variance of the estimate of the treatment effect from the unconditional analysis is rather weak.

Table 2: $\operatorname{var}(\hat{\theta}_U)$ against θ, μ and σ.

		$\operatorname{var}(\hat{\theta}_U)$								
		θ								
		0			2			4		
		μ			μ			μ		
		0	1	2	0	1	2	0	1	2
	0	0.020	0.025	0.048	0.025	0.034	0.068	0.048	0.068	0.152
σ	0.5	0.022	0.027	0.049	0.027	0.035	0.067	0.049	0.067	0.142
	1	0.027	0.033	0.053	0.033	0.040	0.068	0.053	0.068	0.125

6. COMPARISON OF THE EFFICIENCY OF THE CONDITIONAL AND UNCONDITIONAL ANALYSES

The variances of $\hat{\theta}_C$ and $\hat{\theta}_U$, the estimates from the conditional and unconditional analysis respectively, are next compared. The parameter θ is defined in terms of the conditional formulation so that naive estimates of the log odds ratio are not directly comparable. Of the values in Table 3, $\theta = 4$ corresponds to a quite extreme odds ratio.

Table 3: $\mathrm{var}(\hat{\theta}_C)/\mathrm{var}(\hat{\theta}_U)$ against θ, μ and σ.

		$\mathrm{var}(\hat{\theta}_C)/\mathrm{var}(\hat{\theta}_U)$								
		θ								
		0			2			4		
		μ			μ			μ		
		0	1	2	0	1	2	0	1	2
	0	1	1	1	1.54	1.41	1.20	3.76	2.93	1.87
σ	0.5	0.98	0.96	0.93	1.52	1.39	1.18	3.77	3.03	2.00
	1	0.97	0.85	0.76	1.49	1.33	1.08	3.80	3.25	2.30

Table 3 shows the ratio of the variance of the estimate of θ_C to the variance of the estimate of θ_U against θ, μ and σ. The ratio $\mathrm{var}(\hat{\theta}_C)/\mathrm{var}(\hat{\theta}_U)$ is equal to one when $\theta = \sigma = 0$, that is for the null hypothesis with effectively random pairing. As to be expected from the matching, near $\theta = 0$ the gain from using the conditional estimate increases with σ. Especially for larger values of μ and θ, however, the unconditional estimate using the concordant pairs is to be preferred.

The values of the variances as a function of μ, σ and θ were checked by simulation. For θ_C there was good agreement and also for θ_U for small values of σ, but for large σ the calculated variance was larger than the simulated variance.

We now return to testing the hypothesis of no difference between the two groups. In comparing the conditional and unconditional analyses, it is important that the parameters used to specify departures from the null hypothesis have broadly comparable interpretations in the different formulations.

For the conditional analysis, described in Section 6, we take the test statistic to be $T_C = \log(N_{01}/N_{10})$ and in the discussion to follow of the unconditional analysis we take $T_U = \log\{(N_{.1}N_{0.})/(N_{1.}N_{.0})\}$.

Then T_C, interpreted as the logit difference between the two individuals in an arbitrary pair, has asymptotic expected value $E(T_C) = \theta$. At the null hypothesis we have that, asymptotically,

$$\mathrm{var}(T_C) = \left\{ \frac{1}{\frac{1}{2}m_D} + \frac{1}{\frac{1}{2}m_D} \right\} = \frac{4}{m_D} = \frac{4}{m\pi_D},$$

where m is the number of pairs, m_D the number of discordant pairs and π_D the probability of a pair being discordant. The Pitman efficacy (Cox and Hinkley [5, p. 337–338], 1974) for testing the hypothesis that $\theta = 0$ is thus

$$\mathcal{E}_C = \frac{\left\{\partial E(T_C)/\partial\theta\big|_{\theta=0}\right\}^2}{m\mathrm{var}(T_C)\big|_{\theta=0}} = \pi_D.$$

Under the null hypothesis, the probability of a pair being discordant, and hence also $\mathcal{E}_C$, is

$$\pi_D \simeq 2L_0(\mu)L_1(\mu)\left\{1 + \frac{1}{2}\sigma^2\left(1 - 6L_0(\mu)L_1(\mu)\right)\right\}. \qquad (6.1)$$

In the unmatched analysis of Section 7 we have, for instance, that $P(Y_0 = 1) = E\{L_1(A - \theta)\}$. Such probabilities can be calculated approximately by using a Taylor expansion or by approximating $L_1(\cdot)$ by $\Phi(\cdot)$, the standard normal cumulative distribution function. For group 0 this gives

$$P(Y_0 = 1) \simeq L_1(\mu - \theta/2) + \frac{1}{2}\sigma^2 L_1(\mu - \theta/2) L_0(\mu - \theta/2)\left\{L_0(\mu - \theta/2) - L_1(\mu - \theta/2)\right\}.$$

For group 1 an analogous expression holds, with $\mu - \theta/2$ replaced by $\mu + \theta/2$. By a further approximation,

$$\operatorname{logit}\{P(Y_0 = 1)\} \simeq \mu + \frac{1}{2}\sigma^2\left\{L_0(\mu - \theta/2) - L_1(\mu - \theta/2)\right\},$$

so that T_U, the log odds contrast in the unconditional analysis, has asymptotic expected value

$$E(T_U) = \frac{1}{4}\theta + \frac{1}{4}\sigma^2\left\{L_0(\mu + \theta/2) - L_1(\mu + \theta/2) - L_0(\mu - \theta/2) + L_1(\mu - \theta/2)\right\}.$$

Then

$$\frac{\partial E(T_U)}{\partial\theta} \simeq \frac{1}{2}\{1 - \sigma^2 L_0(\mu + \theta/2) L_1(\mu + \theta/2)\}$$

which under the null hypothesis is $\{1 - \sigma^2 L_0(\mu) L_1(\mu)\}/2$. The variance under the null hypothesis is that of the comparison of two independent logits, each based on m observations and thus is

$$\operatorname{var}(T_U) = \frac{1}{2mL_0(\mu)L_1(\mu)}\left\{1 - \frac{1}{2}\sigma^2\left(L_0(\mu) - L_1(\mu)\right)^2\right\},$$

assuming σ^4 is negligible. Therefore the Pitman efficacy for T_U is after some simplification

$$\mathcal{E}_U \simeq \frac{L_0(\mu)L_1(\mu)}{2}\left\{1 + \frac{1}{2}\sigma^2\left(1 - 8L_0(\mu)L_1(\mu)\right)\right\}, \tag{6.2}$$

ignoring terms of order σ^4 and above.

Therefore to assess the relative efficiency for $\theta = 0$, we compare $\mathcal{E}_U$ and $\mathcal{R}_C$. Since in this special case $\mathcal{E}_U$ is smaller than $\mathcal{E}_C$, near the null hypothesis of zero treatment effect the matched design tends to be slightly more efficient than the unmatched one, as is confirmed by the comparison of the variances.

Often $L_0(\mu)L_1(\mu) \simeq 1/4$ and then

$$\mathcal{E}_C \simeq \frac{1}{2}\left(1 - \frac{1}{4}\sigma^2\right)$$

and for comparison

$$\mathcal{E}_U \simeq \frac{1}{2}\left(1 - \frac{1}{2}\sigma^2\right).$$

Thus for testing the hypothesis of no treatment effect the conditional analysis is asymptotically slightly better than the unconditional analysis, depending on the amount of variability between pairs.

7. DISCUSSION

The main qualitative aspects in this discussion, some with broader implications, are as follows. Most importantly, should conclusions be formulated in terms of differences of probabilities or as logistic differences or possibly on some other scale? For any specific set of data the choice is likely to be numerically unimportant if all probabilities are in the central range, say between 0.2 and 0.8 (Cox and Wermuth [7], 1992). The choice becomes important if several sets of data are considered together, when stability of contrasts across data sets, if achievable, is desirable. The direct interpretation of differences of probabilities in terms of the numbers of individuals notionally affected by a change in treatment is attractive but in general decreasingly so at the extremes of the scale, where the logistic comparisons, essentially log ratios at the two ends of the scale, become more appealing, especially so for case-control studies, where there are quite strong specific arguments for the use of logistic differences.

The second general issue applying to the logistic analysis of matched pairs is that the parameter of interest, a difference of log odds, is notionally specific to each pair. This implies, in particular, that it is not directly comparable with the same difference calculated from an unmatched randomized comparison of the same two treatments from the same population. The exception is when the variation between pairs is small. Otherwise some correction based on the inter-pair variability can be made, essentially using the relation between that variability and the proportion of discordant pairs, but such adjustments are likely to be quite fragile.

The third issue is that detailed comparison of the conclusions from different studies, some matched and some totally randomized, requires recognition that different ways of expressing the comparisons of interest by an unknown parameter may be involved.

Finally our detailed results show when the gain in sensitivity from matching is likely to be appreciable.

ACKNOWLEDGMENTS

We are grateful to the referee for constructive comments.

REFERENCES

[1] Agresti, A. (1990). *Categorical Data Analysis*, Wiley, New York.

[2] Aranda-Ordaz, F.J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, **68**, 357–363.

[3] Cox, D.R. (1958a). The regression analysis of binary sequences (with discussion), *J. R. Statist. Soc. B*, **20**, 215–232.

[4] Cox, D.R. (1958b). Two further applications of a model for binary regression, *Biometrika*, **45**, 562–565.

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

[6] Cox, D.R. and Snell, E.J. (1989). *The Analysis of Binary Data*, Second Edition, Chapman and Hall, London.

[7] Cox, D.R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables, *Biometrika*, **79**, 441–461.

[8] Fisher, R.A. (1935). *Design of Experiments*, Oliver and Boyd, Edinburgh.

[9] Lee, Y. (2001). Can we recover information from concordant pairs in binary matched pairs? *J. Applied Statistics*, **28**, 239–246.

[10] McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika*, **12**, 153–157.

[11] Pearson, K. (1907). Reply to certain criticisms of Mr. G.U. Yule, *Biometrika*, **5**, 470–476.

[12] Rasch, G. (1960). *Probability Models for Some Intelligence and Attainment Tests*, Danish Institute for Educational Research, Copenhagen.

[13] Yule, G.U. (1903). Notes on the theory of the association of attributes in statistics, *Biometrika*, **2**, 121–134.