Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 121 (2004) 163–174

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Individual treatment effects in randomized trials with binary outcomes

Gary L. Gadbury[a,*], Hari K. Iyer[b], Jeffrey M. Albert[c]

[a] *Department of Mathematics and Statistics, University of Missouri - Rolla, Rolla, MO 65409, USA*
[b] *Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA*
[c] *Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106-4945, USA*

Received 5 November 2001; accepted 9 December 2002

Abstract

A potential outcomes framework is used to define individual treatment effects in a randomized design comparing two treatments, T and C. When the outcome variable is binary, individual effects may take on one of three values, 0, 1, -1, at any given point in time, but these "individual effects" cannot be measured in practice. Often, in clinical trials, an average effect of the treatment is estimated and a superior treatment is determined from this estimate. However, there may be a proportion of the population that responds favorably to T and another proportion that responds more favorably to C if individual treatment effects vary widely in the population. These proportions are nonidentifiable using data from a two sample completely randomized design, but knowledge regarding their potential magnitude is crucial for assessing the risk involved in administering a treatment to an individual.

We produce identifiable bounds for these proportions using data from an unmatched 2×2 table and then demonstrate the advantages to matching in a matched-pairs design. The advantages hinge on the quality of the matching criteria. We present an extended matched-pairs design that allows estimation of refined bounds. A constructed data example is used to compare the information about individual treatment heterogeneity, and its consequences, that can be gleaned from the different designs.

MSC: 62A01; 62P10; 62K99

Keywords: Clinical trials; Contingency tables; Potential response; Subject-treatment interaction

* Corresponding author. Tel.: +1-573-3414648; fax: +1-573-3414741.
E-mail address: gadburyg@umr.edu (G.L. Gadbury).

0378-3758/$ - see front matter
doi:10.1016/S0378-3758(03)00115-0

1. Introduction

We consider the context of an experimental design (e.g., a clinical trial) that is being used to compare two treatments T and C when the outcome variable is dichotomous. Let p_1 be the proportion of the population responding favorably to T and let p_2 be the proportion of the population responding favorably to C. The issue we consider is that even with $p_1 - p_2 > 0$, there may still be a substantial proportion of "individuals" who would be better off not receiving the treatment T, and this proportion is often "obscured in most of the usual statistical models" (Copas, 1973, p. 468).

Medicine today is geared around statistical information gathered on the basis of population averages and applying it to the individual (Marshall, 1997). Yet, a recent study indicated that there are detrimental effects of treatments on thousands of individuals each year even though the treatments and dosage were considered "appropriate" for the individuals' illnesses (Lazarou et al., 1998). Knowledge of individual variation in response to treatment (subject–treatment interaction) could alert the investigator to the existence of a covariate that, if observed, might help to predict the success of the application of treatment to an individual. We are now entering a phase of scientific development where this "covariate" might be a person's genotype, and studies that search for a gene to explain variation in the perceived effect of a treatment and the development of associated statistical methods have already begun (Elston et al., 1999; Cardon et al., 2000; Rioux, 2000).

In this paper, a potential outcomes framework (Neyman, 1923; Rubin, 1974) is used to build off of some ideas presented in Copas (1973), and the consideration of dichotomous outcomes extends some results by Gadbury and Iyer (2000) who did related work for continuous response variables. We focus on the issues involved in estimating the population proportion, say p^*, that experience a detrimental effect of a treatment T with respect to C. We show that p^* cannot be estimated from observable data but, in unmatched 2×2 tables, bounds for it can be estimated. We then show that more information regarding p^* may be obtained from matched-pairs designs but the ability to estimate p^* hinges on the quality of the matching criteria. Interestingly, subjects need not be perfectly matched in order to obtain valid estimates. We also propose a modification of the paired design that can yield additional information regarding p^*. Throughout, we assume that there is no interference between subjects (Cox, 1958, p. 19) or that Rubin's stable unit treatment value assumption (SUTVA) (Rubin, 1980) holds.

2. Subject-treatment interaction with binary outcomes

Consider a set of two potential outcomes (X, Y), for a subject u in an investigation to compare the effect of a treatment T with respect to a control treatment C. Let X be the outcome of u when exposed to treatment T, and let Y be the outcome if exposed to treatment C. The two outcomes are assumed to be measured at the same moment in time. It is clear that, in practice, only the outcome corresponding to the treatment actually assigned can be observed for a particular subject. Nevertheless,

potential outcomes help to conceptualize a true effect of treatment T with respect to treatment C on u that we define to be $D = X - Y$. When X and Y are dichotomous, the bivariate distribution (X, Y) in an infinite population is assumed to be of the form

(x, y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$P(X = x, Y = y)$	π_1	π_2	π_3	π_4

(1)

where $\sum_i^4 \pi_i = 1$. The distribution of treatment effects, $D = X - Y$, is discrete with D taking on three distinct values, 0, -1, and 1, with probabilities $\pi_1 + \pi_4$, π_2, and π_3, respectively. Suppose that an outcome of 1 indicates an individual is "well" and an outcome of 0 indicates the individual is "ill". At a particular moment in time, $(X, Y) = (1, 0)$ means that $D = 1$ and treatment T caused the individual to be well at that particular moment versus if he/she had received treatment C instead. On the other hand, an unfavorable or detrimental effect of the treatment T with respect to treatment C occurs when $D = -1$. The proportion of the population experiencing a detrimental treatment effect is then π_2 (denoted p^* in the introduction) which is the parameter of interest in this paper. One might also describe the cases where $D = 0$ as unfavorable effects since the treatment T could have other costs involved unrelated to the measured response. Here we simply assume that $D = 0$ indicates cases when the treatment had no effect.

Let $p_1 = E(X) = P(X = 1) = \pi_3 + \pi_4$ and let $p_2 = E(Y) = P(Y = 1) = \pi_2 + \pi_4$, where $E(\cdot)$ is the expectation operator with respect to the population given by (1). The parameters p_1 and p_2 can be estimated from observed data and $p_1 - p_2 = \pi_3 - \pi_2 = E(D)$ is the quantity that is usually used to represent the effect of the treatment T versus treatment C. This quantity represents an *average treatment effect*. The individual parameters, π_3 and π_2, cannot be estimated, and it is the parameter π_2 that quantifies the risk involved in administering the treatment to a particular individual in the population. However, π_2 is bounded by estimable quantities L and U, where

$$\max(0, p_2 - p_1) = L \leqslant \pi_2 \leqslant U = \min(1 - p_1, p_2). \tag{2}$$

Consider data from an unmatched 2×2 table where n_1 subjects were selected to receive treatment T and n_2 subjects were selected to receive C. Suppose there are s_1 "successes" to treatment T and s_2 successes to treatment C. The maximum likelihood estimate (MLE) of p_1 is $\hat{p}_1 = s_1/n_1$, and of p_2, the MLE is $\hat{p}_2 = s_2/n_2$. The MLE for L is $\hat{L} = \max(0, \hat{p}_2 - \hat{p}_1)$ and the MLE for U is $\hat{U} = \min(1 - \hat{p}_1, \hat{p}_2)$. The following example illustrates the use of estimated bounds for π_2.

Example. The data were originally reported in Ezdinli et al. (1976). The data come from a clinical trial comparing two treatments, cytoxan plus prednisone (CP) versus BCNU plus prednisone (BP), for the treatment of lymphocytic lymphoma. The outcome variable measures the response of the tumor in each of 273 patients. The original data included 4 categories of response but, for simplicity, we collapse responses into two categories, a success (the tumor responds favorably to the treatment) or a failure (the tumor does not respond favorably). The data are shown in Table 1 where the number

Table 1
Data on treatment of lymphocytic lymphoma

	Success	Failure	Total
BP	77	61	138
CP	90	45	135

in each cell indicates the number of patients out of 273 that fell into that particular category. We will assume that the 273 individuals in the study are a random sample from the population distribution given by (1). The usual estimates of p_1 and p_2 are $\hat{p}_1 = \frac{90}{135} = 0.67$ and $\hat{p}_2 = \frac{77}{138} = 0.56$, respectively. An estimate of a treatment effect is $\hat{p}_1 - \hat{p}_2 = 0.11$ with estimated standard error 0.059. However, it is the proportion π_3 of the population to which the CP treatment should be applied and the proportion π_2 of the population to which the BP treatment should be applied. We can only estimate that π_2 is between $\hat{L} = 0.0$ and $\hat{U} = 0.33$. Ideally, π_2 would equal zero but we have no way to ascertain this from observed data. The most we can say for this example is that we estimate π_2 to be no larger than 0.33. Next, we consider the role of concomitant information in estimating π_2.

3. Matched 2 × 2 tables

Suppose that additional information is available on subjects in the population and this information is used to match subjects into pairs. We continue to focus on π_2, and we parameterize matching criteria to show what conditions must be met for π_2 to be estimable. A parameterization of matching criteria for continuous outcome variables in a finite population can be found in Gadbury (2001).

Assume an infinite population of the form given by (1), and that a selected individual from this population is matched to another individual on the basis of covariate(s) or other subjective information. Subjects within a pair then receive labels of subjects 1 and 2 from a random coin toss, so subject 1 has potential outcomes (X_1, Y_1) and 2 has potential outcomes (X_2, Y_2). The joint distribution of potential outcomes is given by $P[(x_1, y_1), (x_2, y_2)]$. Subjects within pairs are exchangeable (due to the random labeling), so $P[(x_1, y_1), (x_2, y_2)] = P[(x_2, y_2), (x_1, y_1)]$. Let ψ_{ij}, $i,\ j = 1, 2, 3, 4$, denote the joint probabilities of potential outcomes for a pair so that the population distribution of potential outcomes has a form given by the following array:

	(x_2, y_2) $(0,0)$	(x_2, y_2) $(0,1)$	(x_2, y_2) $(1,0)$	(x_2, y_2) $(1,1)$
$(x_1, y_1) = (0,0)$	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}
$(x_1, y_1) = (0,1)$	ψ_{21}	ψ_{22}	ψ_{23}	ψ_{24}
$(x_1, y_1) = (1,0)$	ψ_{31}	ψ_{32}	ψ_{33}	ψ_{34}
$(x_1, y_1) = (1,1)$	ψ_{41}	ψ_{42}	ψ_{43}	ψ_{44}.

(3)

Marginal probabilities are row totals, $\pi_i = \sum_j \psi_{ij}$ and column totals $\pi_j = \sum_i \psi_{ij}$ for $i, j = 1, 2, 3, 4$. By exchangeability, $\psi_{ij} = \psi_{ji}$, and in results that follow we use the term

ψ_{ij} with $i \leqslant j$. When matching is perfect, $\sum_i \psi_{ii} = 1$, so the parameters ψ_{ij} have an interpretation that reflects the quality of pairing that could be done in the sampled population using available covariate information.

We assume that treatments T and C are randomly assigned within a pair with equal probabilities. Without loss of generality (due to random labeling of subjects within a pair), we assume that, for each pair, subject 1 receives T and subject 2 receives C. Observed data are of the form (X_1, Y_2), and the following 4 outcomes occur with probabilities $\theta_i,\ i = 1,\ldots,4$ that, after some algebra, can be written as

$$\begin{aligned}
\theta_1 &= P[X_1 = 0, Y_2 = 0] = \pi_1 - \delta,\\
\theta_2 &= P[X_1 = 0, Y_2 = 1] = \pi_2 + \delta,\\
\theta_3 &= P[X_1 = 1, Y_2 = 0] = \pi_3 + \delta,\\
\theta_4 &= P[X_1 = 1, Y_2 = 1] = \pi_4 - \delta,
\end{aligned} \tag{4}$$

where $\delta = \psi_{14} - \psi_{23}$.

Suppose that n pairs are selected at random from the population in (3) and treatments are assigned at random within pairs as described above. Observed data for the n pairs are as follows:

X_1	Y_2	Frequency
0	0	s_1
0	1	s_2
1	0	s_3
1	1	s_4

where $\sum_{i=1}^{4} s_i = n$. The MLE for $\theta_i,\ i = 1,\ldots,4$ is given by s_i/n which is also unbiased. The next proposition follows immediately from these results, and a following definition helps to summarize the key points thus far.

Proposition 1. *For a population structure of the form given by the expression* (1) *and for* $i = 1,\ldots,4$, s_i/n *is an unbiased estimator of* π_i *if and only if* $\psi_{14} = \psi_{23}$.

Definition 1. Let the potential outcome for a pair be $\{(x_1, y_1), (x_2, y_2)\}$ which is equivalent to $\{(x_2, y_2), (x_1, y_1)\}$. The pair is considered to be:

- a perfect match if $x_1 = x_2$ and $y_1 = y_2$,
- a treatment mismatch if $x_1 \neq x_2$ and $y_1 = y_2$,
- a control mismatch if $x_1 = x_2$ and $y_1 \neq y_2$, and
- a double mismatch if $x_1 \neq x_2$ and $y_1 \neq y_2$.

Four points are worth noting. First, the average treatment effect is $\theta_3 - \theta_2 = \pi_3 - \pi_2$. The random sample and random treatment assignment assure that this can be estimated regardless of the matching criteria. Second, there could be two occurrences of a double mismatch, $\{(x_1, y_1), (x_2, y_2)\} = \{(0,0), (1,1)\}$ or $\{(x_1, y_1), (x_2, y_2)\} = \{(0,1), (1,0)\}$. The condition in Proposition 1 for unbiased estimation of $\pi_i,\ i = 1,\ldots,4$, is that the

probabilities of these two occurrences are equal. Third, a stronger condition would be that the probability of a double mismatch is zero, and this condition is sufficient for unbiased estimation of the π_i. This condition would imply that $\psi_{14}=\psi_{23}=0$. A positive probability of selecting a pair with only a treatment mismatch or a control mismatch does not affect the ability to obtain an unbiased estimator for the π_i. Finally, an even stronger condition would be that the probability of a perfect match is equal to 1. That is, $\sum_i \psi_{ii} = 1$, which implies that only perfectly matched pairs were present in the sampled population.

As the matching improves, δ in Eqs. (4) approaches zero. When δ is not known, one can obtain MLEs for certain bounds of π_2. These bounds are given in the following proposition.

Proposition 2. *Bounds for* π_2, *using the matched-pairs design*, *are* L *and* U *where*,

$$L = \max(0, \theta_2 - \theta_3), \quad U = \min(\theta_2 + \theta_1, \theta_2 + \theta_4). \tag{5}$$

These bounds are the same bounds given by Eq. (2), *derived for the unmatched design. Furthermore*, *there exists a distribution of the form given by* (3) *for which the bounds are attained. Finally*, L *and* U *are identifiable and have MLEs*,

$$\hat{L} = \max(0, \hat{\theta}_2 - \hat{\theta}_3), \quad \hat{U} = \min(\hat{\theta}_2 + \hat{\theta}_1, \hat{\theta}_2 + \hat{\theta}_4), \tag{6}$$

where $\hat{\theta}_i = s_i/n$.

Proof. Identifiability of the bounds follows from the fact that there is a unique MLE for θ_i, $i=1,\ldots,4$. For the upper bound, $\pi_1+\pi_2=\theta_1+\theta_2$ and $\pi_2+\pi_4=\theta_2+\theta_4$ imply that $\pi_2 \leqslant \min(\theta_1+\theta_2, \theta_2+\theta_4)$. Furthermore, π_2 will equal U ($\theta_1+\theta_2$ if $\theta_1 \leqslant \theta_4$ and $\theta_2+\theta_4$ if $\theta_4 \leqslant \theta_1$) if the joint distribution in (3) has ψ_{ij} given by, respectively,

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	0	0	0	0
(0,1)	0	θ_2	θ_1	0
(1,0)	0	θ_1	θ_3	0
(1,1)	0	0	0	$\theta_4-\theta_1$

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	$\theta_1-\theta_4$	0	0	0
(0,1)	0	θ_2	θ_4	0
(1,0)	0	θ_4	θ_3	0
(1,1)	0	0	0	0.

(7)

Similarly, $\pi_3-\pi_2=\theta_3-\theta_2$ produces the lower bound for π_2. The bound is attained for the following two distributions, depending on whether $\theta_2 \leqslant \theta_3$ or not,

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	θ_1	0	0	θ_2
(0,1)	0	0	0	0
(1,0)	0	0	$\theta_3-\theta_2$	0
(1,1)	θ_2	0	0	θ_4

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	θ_1	0	0	θ_3
(0,1)	0	$\theta_2-\theta_3$	0	0
(1,0)	0	0	0	0
(1,1)	θ_3	0	0	θ_4.

(8)

Finally, matching subjects into pairs should not worsen one's ability to bound π_2 since $\theta_2+\theta_1=1-p_1$ and $\theta_2+\theta_4=p_2$, and $\min(1-p_1, p_2)$ was the upper bound derived for the unmatched design in the prior section. The lower bound for the unmatched design is equivalent to that in (5). □

4. An extended matched-pairs design

A refined set of bounds for π_2 can be derived that depends on the quality of matching, and this quality of matching may be estimable in an extended matched-pairs design.

Proposition 3. *Let* $p_1=P(X=1)$ *and* $p_2=P(Y=1)$ *as given in Section* 2.

$$|\delta| \leqslant \min\{p_1 - P(X_1=1,X_2=1),\, p_2 - P(Y_1=1,Y_2=1)\}.$$

So $\pi_2 \in [L^*, U^*]$ *where,*

$$L^* = \max\{0, \theta_2 - \min\{p_1 - P(X_1=1,X_2=1),\, p_2 - P(Y_1=1,Y_2=1)\}\},$$

$$U^* = \theta_2 + \min\{p_1 - P(X_1=1,X_2=1),\, p_2 - P(Y_1=1,Y_2=1)\}$$

and where $\theta_2 = P[X_1=0, Y_2=1]$.

Proof. $|\delta| \leqslant \psi_{14}+\psi_{23}$ which is the $\frac{1}{2}$ of the probability of a double mismatch.

$$\begin{aligned}\psi_{14}+\psi_{23} &= (\pi_3+\pi_4) - (\psi_{13}+\psi_{24}+2\psi_{34}+\psi_{33}+\psi_{44}) \\ &\leqslant (\pi_3+\pi_4) - (2\psi_{34}+\psi_{33}+\psi_{44}) = p_1 - P(X_1=1,X_2=1).\end{aligned}$$

Also,

$$\begin{aligned}\psi_{14}+\psi_{23} &= (\pi_2+\pi_4) - (\psi_{12}+\psi_{22}+2\psi_{24}+\psi_{34}+\psi_{44}) \\ &\leqslant (\pi_2+\pi_4) - (\psi_{22}+2\psi_{24}+\psi_{44}) = p_2 - P(Y_1=1,Y_2=1).\end{aligned}$$

The above shows $|\delta| \leqslant \min\{p_1 - P(X_1=1,X_2=1),\, p_2 - P(Y_1=1,Y_2=1)\}$, and bounds for π_2 follow from the expressions in (4). □

The symmetry of the joint distribution in (3) results in the following relationships that are not too difficult to show:

$$p_1 - P(X_1=1,X_2=1) = 1 - p_1 - P(X_1=0,X_2=0),$$

$$p_2 - P(Y_1=1,Y_2=1) = 1 - p_2 - P(Y_1=0,Y_2=0).$$

The two quantities, $P(X_1=1,X_2=1)$ and $P(Y_1=1,Y_2=1)$ (or $P(X_1=0,X_2=0)$ and $P(Y_1=0,Y_2=0)$), could not be estimated in the standard matched-pairs design, but they can be estimated using an extended design.

Suppose that N pairs are a random sample from the population of potential outcomes in (3). Assume that for n_1 pairs the treatment assignment is (T,T) (i.e., both subjects receive treatment T), for n_2 pairs it is (C,C), and for n_3 pairs T and C are randomly assigned (again we assume subject 1 receives T), and $n_1+n_2+n_3=N$.

For the ith subject in a pair, $i=1,2$, let X_i be the outcome if treatment T was received and Y_i be the outcome if C was received. Define the following 8 probabilities:

$$\begin{aligned}
&\alpha_1 = P[X_1 = 0, X_2 = 0], \quad \beta_1 = P[Y_1 = 0, Y_2 = 0],\\
&\alpha_2 = P[X_1 = 0, X_2 = 1], \quad \beta_2 = P[Y_1 = 0, Y_2 = 1],\\
&\alpha_3 = P[X_1 = 1, X_2 = 0], \quad \beta_3 = P[Y_1 = 1, Y_2 = 0],\\
&\alpha_4 = P[X_1 = 1, X_2 = 1], \quad \beta_4 = P[Y_1 = 1, Y_2 = 1].
\end{aligned} \tag{9}$$

By exchangeability $\alpha_2=\alpha_3$ and $\beta_2=\beta_3$, so $2\alpha_2$ and $2\beta_2$ are the probabilities of mismatching subjects on the treatment variable and control variable, respectively. Moreover, α_4 and β_4 (or α_1 and β_1) are probabilities that must be estimated to apply Proposition 3. Observed data are of the following form:

X_1	X_2	Freq.
0	0	a_1
0	1	a_2
1	0	a_3
1	1	a_4

Y_1	Y_2	Freq.
0	0	b_1
0	1	b_2
1	0	b_3
1	1	b_4

X_1	Y_2	Freq.
0	0	s_1
0	1	s_2
1	0	s_3
1	1	s_4

where, $\sum_i a_i = n_1$, $\sum_i b_i = n_2$, and $\sum_i s_i = n_3$. With $\theta_1,\ldots,\theta_4$ as given in Eqs. (4), the likelihood, $\mathscr{L}(\cdot)$, of observed data has the following form.

$$\begin{aligned}
\mathscr{L}(\cdot) \propto\ & \alpha_1^{a_1}\alpha_2^{(a_2+a_3)}\alpha_4^{(n_1-a_1-a_2-a_3)}\beta_1^{b_1}\beta_2^{(b_2+b_3)}\beta_4^{(n_2-b_1-b_2-b_3)}\\
& \theta_1^{s_1}\theta_2^{s_2}\theta_3^{s_3}(1-\theta_1-\theta_2-\theta_3)^{n_3-s_1-s_2-s_3}.
\end{aligned} \tag{10}$$

There are constraints on the parameters in (10). They are,

$$\alpha_1 + 2\alpha_2 + \alpha_4 = 1, \quad \beta_1 + 2\beta_2 + \beta_4 = 1,$$

$$\sum \theta_i = 1, \quad \alpha_i, \beta_i, \theta_i \in [0,1],\ i = 1,2,3,4.$$

The likelihood expression in (10) contains 5 distinct parameters since $\alpha_2 = \theta_1 + \theta_2 - \alpha_1$ and $\beta_2 = \theta_1 + \theta_3 - \beta_1$. Solving for the MLEs involves solving a system of 5 nonlinear equations subject to the constraints. The equations can be reparameterized into a system of 5 nonlinear equations with no constraints, but their solution still requires iterative numerical methods. The results from one method are illustrated in the next section.

The interpretation of the additional information from the new design is apparent. The parameter α_2 is one half the probability of obtaining a pair that did not match on the treatment variable X, and β_2 is one half the probability of obtaining a pair that is mismatched on the Y variable. If matching is good, one would expect that one or both of α_2 and β_2 are small. To obtain a good bound for π_2 it is not necessary to have good matching on both variables X and Y, just good matching on one or the other is sufficient. Furthermore, if one believes that matching is good, then the design could be

simplified by only including one set of pairs for which both subjects receive the same treatment.

5. A constructed example for illustration

The following illustration compares the information that can be obtained from the various designs discussed above. A population distribution of the form given by (3) is used to generate the data. The joint probability values, i.e., ψ_{ij}, have been multiplied by 16 for easier readability.

	(x_2, y_2) $(0,0)$	(x_2, y_2) $(0,1)$	(x_2, y_2) $(1,0)$	(x_2, y_2) $(1,1)$
$(x_1, y_1) = (0,0)$	2.00	0.75	1.00	0.25
$(x_1, y_1) = (0,1)$	0.75	0.50	0.50	0.25
$(x_1, y_1) = (1,0)$	1.00	0.50	3.50	1.00
$(x_1, y_1) = (1,1)$	0.25	0.25	1.00	2.50.

(11)

The marginal probabilities are $(\pi_1, \pi_2, \pi_3, \pi_4) = (\frac{1}{16})(4, 2, 6, 4) = (0.250, 0.125, 0.375, 0.250)$. The average treatment effect is $\pi_3 - \pi_2 = 0.250$. Values of parameters that were presented for the unmatched design are $P(X = 1) = p_1 = 0.625$, $P(Y = 1) = p_2 = 0.375$, and the lower and upper bounds for π_2 (Eq. (2)) are 0 and 0.375, respectively.

Values of parameters for the matched-pairs designs are
$(\theta_1, \theta_2, \theta_3, \theta_4) = (0.266, 0.109, 0.359, 0.266)$, and $\delta = -0.25/16 = -0.0156$.

A random sample of 400 "matched" observations was generated from the population given by (11). From these, $n_1 = 50$ pairs were randomly selected to receive (T, T), $n_2 = 50$ pairs were randomly selected to receive (C, C), and subjects in the remaining $n_3 = 300$ pairs received different treatments. These sample sizes were selected for illustration but the choice of n_1, n_2, and n_3 is a nontrivial design issue that is briefly discussed in the closing section.

We first assume that the 50 matched pairs for which X is observed for both subjects are treated as two independent subjects receiving treatment T, and similarly for the 50 pairs for which Y is observed. Thus, we essentially ignore the matching. A response of 1 is taken to be a "success". The resulting "unmatched" 2×2 table of observed data is

	Success	Failure	Total
Treatment	265	135	400
Control	151	249	400

From these data, $\hat{p}_1 - \hat{p}_2 = 0.6625 - 0.3775 = 0.2850$ which is the estimated mean treatment effect. The estimated lower bound for π_2 is 0 and the estimated upper bound is $1 - \hat{p}_1 = 0.3373$, an estimate of the actual upper bound of 0.375.

Now we consider the added information and interpretation that is available by considering a simple matched-pairs design for which one subject in a pair receives treatment T with the other subject receiving treatment C. There are 300 such pairs and the observed data are

X_1	Y_2	Frequency
0	0	77
0	1	26
1	0	116
1	1	81

The multinomial parameter estimates are $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) = (0.2567, 0.0867, 0.3867. 0.2700)$. If the probability of a double mismatch in the paired population was small, then one might believe that $\hat{\theta}_2 = 0.0867$ would be close to π_2 within a specified margin of error (recall from above that $\pi_2 = 0.1250$ for this example). The benefit of the paired design depends on how well subjects were matched. With no added information regarding the quality of matching, the most that can be done is to, again, estimate lower and upper bounds for π_2. The estimated lower bound is 0, and the estimated upper bound is 0.3433 which is not any improvement since it estimates the same bound as for the unmatched design. However, one has the opportunity to make δ small by "good" matching so that θ_2 is close to π_2. This feature of the matched-pairs design was not available in the unmatched design.

Finally we consider the added information available in the 100 pairs in which both subjects received the same treatment. These data look like,

X_1	X_2	Freq.
0	0	10
0	1	2
1	0	10
1	1	28

Y_1	Y_2	Freq.
0	0	21
0	1	8
1	0	6
1	1	15

The numerical routine "proc NLIN" in SAS was used to compute the MLEs for the parameters in Eq. (10). They are,

$$(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4) = (0.2476, 0.0894, 0.3794, 0.2835),$$

$$(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4) = (0.2151, 0.1220, 0.1220, 0.5410),$$

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4) = (0.4937, 0.1333, 0.1333, 0.2397).$$

Other estimates are $\hat{p}_1 = 0.6630$ and $\hat{p}_2 = 0.3730$. The estimated lower and upper bounds for π_2 are 0 and 0.2114, respectively, so the extended design produced a tighter upper bound for π_2.

6. Discussion and future work

This paper produced bounds for π_2, and it highlighted the fact that design considerations can facilitate tighter bounds that are estimable. Future work will develop the properties of estimators thereby making interval estimators possible. The design issue of selecting $n_i, i = 1, 2, 3$, referred to in the previous section, is an important consideration. For instance, selecting pairs for which both subjects receive the same treatment may improve the ability to estimate a tight bound for π_2, but it is likely to do so at the expense of some loss of efficiency in estimating a mean treatment effect. These tradeoffs may become more apparent as more is learned about the properties of the MLEs for the bounds.

Other designs are also being investigated such as block designs that may produce yet tighter bounds for π_2. Since the bounds often depend on the quality of pairing or quality of the blocking variable, Bayesian approaches that incorporate prior information regarding blocking or matching criteria might be useful.

The issue of estimating π_2 revolved around "unexplained" individual treatment variation in matched and unmatched designs. This parameter could also provide information on the "possible" magnitude of a treatment by covariate interaction, thus revealing the possibility of explainable variation. If a candidate covariate is for instance, a marker genotype, then knowledge of a potentially large interaction with treatment can be useful before conducting expensive assays.

Finally, since dichotomous responses might reflect survival (or death) at a certain point in time, a subject of further research is to extend ideas in this paper and in Gadbury and Iyer (2000) to censored survival data.

Acknowledgements

We thank the editor, an editorial board, and two anonymous referees for helpful comments that improved the manuscript.

References

Cardon, L.R., Idury, R.M., Harris, T.J.R., et al., 2000. Testing drug response in the presence of genetic information: sampling issues for clinical trials. Pharmacogenetics 10, 503–510.

Copas, J.B., 1973. Randomization models for the matched and unmatched 2 × 2 tables. Biometrika 60, 467–476.

Cox, D.R., 1958. Planning of Experiments. Wiley, New York.

Elston, R.C., Idury, R.M., Cardon, L.R., Lichter, J.B., 1999. The study of candidate genes in drug trials: sample size considerations. Statist. Med. 18, 741–751.

Ezdinli, E., Pocock, S., Berard, C.W., et al., 1976. Comparison of intensive versus moderate chemotherapy of lymphocytic lymphomas: a progress report. Cancer 38, 1060–1068.

Gadbury, G.L., 2001. Randomization inference and bias of standard errors. Amer. Statist. 55, 310–313.

Gadbury, G.L., Iyer, H.K., 2000. Unit-treatment interaction and its practical consequences. Biometrics 56, 882–885.

Lazarou, J., Pomeranz, B.H., Cory, N., 1998. Incidence of adverse drug reactions in hospitalized patients. J. Amer. Med. Assoc. 279, 1200–1205.

Marshall, A., 1997. Laying the foundations for personalized medicines. Nat. Biotechnol. 15, 955–957.
Neyman, J., 1923. On the application of probability theory to agricultural experiments. Essay on principles. Roczniki Nauk Rolniczych X, 1–51 (in Polish) (English translation of Section 9 by D.M. Dabrowska, T.P. Speed, Statist. Sci. 5, 465–480).
Rioux, P.P., 2000. Clinical trial in pharacogenetics and pharmacogenomics: methods and applications. Amer. J. Health—Syst. Pharm. 57, 887–898.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psych. 66, 688–701.
Rubin, D.B., 1980. Comment on Basu's randomization analysis of experimental data. J. Amer. Statist. Assoc. 75, 591–593.