

RE: "POSSIBILITY OF SELECTION BIAS IN MATCHED CASE-CONTROL STUDIES
USING FRIEND CONTROLS"

Flanders and Austin (1) demonstrate that the rate ratio from a matched case-control study using friend controls can be biased and state that this bias will arise when "the study exposure is a determinant of friendship" (1, p. 150). This bias, in fact, results rather from the fact that case and friend-control are commonly *not* in a reciprocal relationship, so that a certain subject (e.g., a particularly "gregarious" person) will be the potential friend-control for two or more subjects, while another (e.g., a particularly "asocial" person) will tend to be no one's friend-control. As a consequence, gregarious subjects will be overrepresented among the controls, and thus, if the exposure rate among gregarious subjects differs from that among asocial subjects, bias will result.

Consider, however, a case-control design which arises from subdividing the population at risk into mutually exclusive subpopulations of size 2, such that should one member of a subpopulation become a case, the other member becomes his control—a reciprocal matched-pair design. Bias does not arise in a reciprocal design, *even if the subdivision of the population* (and hence of the case-control pairs) *is made on the basis of friendship* determined partly (or solely) by the study exposure.

In such a design, the parameters of Flanders and Austin (1) p , d_1 , and d_2 are functionally not independent, so that once any two are given the third is fixed, and this fixed third value ensures that the rate ratio is correctly estimated. (Note: In what follows we have used their notation except for writing ψ as R and δ as d .)

The nonindependence of p , d_1 , and d_2 can be demonstrated as follows:
The population and cases are

	<u>Population</u>	<u>Cases</u>
Exposed (E)	Np	$NpRI_0$
Unexposed (U)	$N(1 - p)$	$N(1 - p)I_0$

Now suppose that the population is broken down into "reciprocal sets of size 2" and write

<u>Reciprocal set</u>	<u>No. of sets in the population</u>	<u>Cases</u>
E, E	F_1	$2F_1RI_0$
E, U	F_2	$F_2RI_0 + F_2I_0$
U, U	F_3	$2F_3I_0$

For this population, we then have

$$N = 2(F_1 + F_2 + F_3) \quad (1)$$

and

$$p = (2F_1 + F_2)/N. \quad (2)$$

The probability that a case who is E will have a control who is E is

$$p + d_1 = 2F_1/(2F_1 + F_2). \quad (3)$$

Equations 1, 2, and 3 for fixed N , p , and d_1 , define F_1 , F_2 , and F_3 . Thus, the probability that a case who is U will have a control who is U , i.e.,

$$1 - p + d_2 = 2F_3/(2F_3 + F_2) \quad (4)$$

is fixed, so that d_2 is fixed. The rate ratio estimate is biased by a factor

$$B = p(1 - p - d_1)/[(1 - p)(p - d_2)] \quad (5)$$

(This is Flanders and Austin's equation 1.) Substituting equations 3 and 4 into equation 5, we find

$$B = p(2F_3 + F_2)/[(1 - p)(2F_1 + F_2)] = 1.$$

Thus, there is no bias, and, in particular, this lack of bias holds true even when exposed subjects (i.e., population members) tend to form reciprocal sets with other exposed subjects and unexposed subjects tend to form reciprocal sets with other unexposed subjects (so that F_2 is small compared with F_1 and F_3).

Reciprocal designs can be extended to 1 to k matched case-control studies by dividing the population into mutually exclusive subpopulations of size $(k + 1)$ such that if a member of a subpopulation becomes a case, the members of the same subpopulation who are still at risk become the case's matched controls. Such designs are also unbiased even if the subpopulations are defined on the basis of friendship, since they can be viewed as special cases of the case-control design in which each case is matched to one or more controls randomly selected from population members at risk and in the same stratum (subpopulation) as the case (such designs are known to be valid (2)). The special case is that in which each of the mutually exclusive subpopulations of size $(k + 1)$ defines a unique stratum, and a control sampling fraction of 1 is applied within strata. We discuss these ideas in more detail elsewhere (3).

In summary, a valid design requires that each case and matched control(s) belong to the same subpopulation of the underlying cohort study. The usual methods of selecting friend-controls almost always involve some form of asymmetry (nonreciprocity) of case and control(s), so that it is not possible to regard them as arising from the same subpopulation in an underlying cohort study. Friend-controls are thus generally to be avoided. If, however, "friendship groups" can be defined in such a way that they make up a mutually exclusive and exhaustive subdivision of the population at risk (e.g., sixth grade class membership), then this will be a perfectly valid basis of a case-control design, although such a design is likely to suffer from overmatching, which will reduce its efficiency.

REFERENCES

1. Flanders WD, Austin H. Possibility of selection bias in matched case-control studies using friend controls. *Am J Epidemiol* 1986;124:150-3.
2. Breslow NE, Day NE. The analysis of case-control studies. Lyon, France: International Agency for Research on Cancer scientific publications no. 32, 1980:69-73.
3. Robins J, Pike MC. The validity of case-control studies with nonrandom selection of controls. (unpublished manuscript).

M. C. Pike
Department of Preventive Medicine

University of Southern California
School of Medicine
Los Angeles, CA 90033

James Robins
Occupational Health Program and
Department of Biostatistics
Harvard University School of Public Health
Boston, MA 02115