MATCHING AFTER RANDOM ASSIGNMENT: IMPUTING PERFORMANCE[1]

George Cave, Manpower Demonstration Research Corporation
Three Park Avenue, New York, New York 10016

I. Introduction

Random assignment of subjects to program treatment or to a control group eliminates many problems in evaluation research. However, some problems remain even after random assignment has taken place. In particular, there is the problem of possible "noncompliance bias" (see Cave, 1988): when program participation or other performance measures are not the same for all subjects assigned to treatment, the average impact of the program on treatment assignees may misrepresent quite seriously the program impact on some performance subgroups. It is especially troubling when the true program effect on some performance subgroups is zero, and the average impact on treatment assignees is driven by much higher than average impacts on other performance subgroups. Such impact heterogeneity would not present an insurmountable problem if (1) it were possible to measure outcomes with and without the program for those program assignees in each performance subgroup; or (2) the performance of control assignees could be predicted using the performance of program assignees, allowing treatment-control differences to be stratified by performance; or (3) identifying restrictions could be imposed to allow the parameters of interest to be estimated. While alternative (1) is not available, there have been several past attempts to implement alternatives (2) and (3).[2] This paper develops two new approaches. One approach is to "uncollapse" contingency tables for outcomes by treatment assignment into subtables for each performance category. The other approach is to predict control performance using pair matching on pre-assignment covariates. The techniques are applied to data from a recent wage subsidy program evaluation.[3]

II. The Basic Problem

In its simplest form, the problem of stratifying impacts by performance involves classifying members of a target population by values of three dummy variables. If G represents the group to which each person may be assigned--the control group (C), or the group eligible to receive treatment (T); if Q represents a potential performance category (such as "would be placed in a subsidized job," observed only for those assigned to treatment); and if Y represents a categorical post-program outcome (such as earnings above or below some fixed target amount, such as $2080 per calendar quarter); then there are 2 x 2 x 2 = 8 elementary population frequencies $P_{QY}{}^{G}$. These parameters appear in Table 1, which gives expected sample category counts $m_{QY}{}^{G}$ when $n^T = m_{++}{}^T$ are assigned at random to the treatment, and $n^C = m_{++}{}^C$ are assigned at random to the control group.

Table 1
Expected Category Counts, by Assignment, Performance, and Outcome

Assignment Group (G)	Performance Indicator (Q)	Outcome Group: Earnings Target Not Reached (Y=0)	Outcome Group: Target Reached(Y=1)	Row Margins
Treatment(G=T)	NotPlaced(0)	$m_{00}{}^T = n^T P_{00}{}^T$	$m_{01}{}^T = n^T P_{01}{}^T$	$m_{0+}{}^T$
	Placed (Q=1)	$m_{10}{}^T = n^T P_{10}{}^T$	$m_{11}{}^T = n^T P_{11}{}^T$	$m_{1+}{}^T$
	Column Margins	$m_{+0}{}^T = n^T P_{+0}{}^T$	$m_{+1}{}^T = n^T P_{+1}{}^T$	n^T
Control(G=C)	NotPlaced(0)	$m_{00}{}^C = n^C P_{00}{}^C$	$m_{01}{}^C = n^C P_{01}{}^C$	$m_{0+}{}^C$
	Placed (Q=1)	$m_{10}{}^C = n^C P_{10}{}^C$	$m_{11}{}^C = n^C P_{11}{}^C$	$m_{1+}{}^C$
	Column Margins	$m_{+0}{}^C = n^C P_{+0}{}^C$	$m_{+1}{}^C = n^C P_{+1}{}^C$	n^C

A measure of the effectiveness of assignment to the program is a log odds ratio which uses column marginal probabilities,

$$(1)\quad \delta_{+} = \log(\frac{P^T_{+1}}{P^T_{+0}}) - \log(\frac{P^C_{+1}}{P^C_{+0}}) = \log\{(\frac{P^T_{+1}}{P^T_{+0}})/(\frac{P^C_{+1}}{P^C_{+0}})\}$$
$$= \log(\frac{S^T_{+}}{S^C_{+}}) = \log\{(\frac{m^T_{+1}}{m^T_{+0}})/(\frac{m^C_{+1}}{m^C_{+0}})\}.$$

This parameter is a measure of the difference in the odds of success for those assigned to the treatment ($S_+{}^T$), relative to the odds of success for those assigned to the control group ($S_+{}^C$).

One problem in using (1) to gauge the effectiveness of the program is that it measures only the effect of assignment to placement in a subsidized job, not the effect of actual placement in a subsidized job. The effect of placement in a subsidized job is

$$(2)\quad \delta_{1} = \log(\frac{P^T_{11}}{P^T_{10}}) - \log(\frac{P^C_{11}}{P^C_{10}}) = \log\{(\frac{P^T_{11}}{P^T_{10}})/(\frac{P^C_{11}}{P^C_{10}})\}$$
$$= \log(\frac{S^T_{1}}{S^C_{1}}) = \log\{(\frac{m^T_{11}}{m^T_{10}})/(\frac{m^C_{11}}{m^C_{10}})\}.$$

For those members of the target population for the program who would be placed if selected for the program, this parameter is a measure of the difference in the odds of success when assigned to the treatment, relative to the odds of success when assigned to the control group.

For the rest of the target population, those who would not be placed if selected for the program, a measure of the difference in the odds of success when assigned to the treatment, relative to the odds of success when assigned to the control group is

$$(3)\quad \delta_{0} = \log(\frac{P^T_{01}}{P^T_{00}}) - \log(\frac{P^C_{01}}{P^C_{00}}) = \log\{(\frac{P^T_{01}}{P^T_{00}})/(\frac{P^C_{01}}{P^C_{00}})\}$$
$$= \log(\frac{S^T_{0}}{S^C_{0}}) = \log\{(\frac{m^T_{01}}{m^T_{00}})/(\frac{m^C_{01}}{m^C_{00}})\}.$$

A program can have a substantial effect on those who do not complete it. In the case of a wage subsidy program, those who are not placed in subsidized jobs still might be affected by contacts with job developers. One reason a sample member might not have taken a subsidized job might be that a job developer showed her how to improve her job applications and on her own she found an employer who did not need to be told about the subsidy. On the other hand, some sample members might have their employment prospects worsened, at least temporarily, by having job developers raise their wage expectations in vain. Thus research hypotheses to be tested in evaluating a wage subsidy program include (1) the effect of the program on those who are placed is not the same as the effect of the program on those who are not placed; and (2) the effect of the program on those who are not placed is not zero. In terms of the parameters just defined, these hypotheses give rise to the following null hypotheses:

$(H1)\ \delta_1 = \delta_0;$

$(H2)\ \delta_0 = 0.$

If H2 cannot be rejected, the case is strengthened for attributing all of the program impact to the other performance category. See Cave (1988) for an estimator to be used in this situation.

Table 2 shows that the basic problem with this approach is that outcomes stratified by potential performance are not observed for those assigned to the control group. Only the column marginals are available for controls. However, sample analogues for all the probabilities applicable to those assigned to the treatment are observed.

Table 2
Observed Category Counts, by Assignment, Performance, and Outcome

Assignment Group (G)	Performance Indicator (Q)	Outcome Group: Earnings Target Not Reached (Y=0)	Outcome Group: Target Reached(Y=1)	Row Margins
Treatment(G=T)	NotPlaced(0)	$x_{00}{}^T = n^T\ p_{00}{}^T$	$x_{01}{}^T = n^T p_{01}{}^T$	$x_{0+}{}^T$
	Placed (Q=1)	$x_{10}{}^T = n^T\ p_{10}{}^T$	$x_{11}{}^T = n^T p_{11}{}^T$	$x_{1+}{}^T$
	Column Margins	$x_{+0}{}^T = n^T\ p_{+0}{}^T$	$x_{+1}{}^T = n^T p_{+1}{}^T$	n^T
Control(G=C)	NotPlaced(0)	---a	---a	---a
	Placed (Q=1)	---a	---a	---a
	Column Margins	$x_{+0}{}^C = n^C\ p_{+0}{}^C$	$x_{+1}{}^C = n^C p_{+1}{}^C$	n^C

NOTE: aNot observed.

There is no ideal and simple solution to the problem posed by the unobserved cells. However, this paper proposes two approaches. One approach, explained in the next section, is to borrow restrictions from the top half of the table to identify the missing cells in the bottom half of the table. Matching is another approach, as explained in Section IV. From the point of view of matching, the basic problem here is that performance is observed for members of the treatment group, but not for their counterparts in the control group. When treatment and control groups have been constructed using random assignment, it seems reasonable that each performance subgroup in the treatment group has a counterpart in the control group. The two groups need only be matched together somehow in order to fill in the blanks in Table 2.

III. "Uncollapsing" a Partially Categorized Contingency Table

The problem posed by the unobserved cells in Table 2 is serious but not hopeless. Since random assignment makes it reasonable that each performance subgroup in the treatment group should have a counterpart in the control group, it makes sense to borrow the row margin relative frequencies from the top half of the table and give them to the bottom half. That is, $x_{0+}{}^C$, the unobserved number of controls who would not have been placed had their random assignment gone the other way, should be replaced with $n^C p_{0+}{}^T$, and similarly $x_{1+}{}^C$ should be replaced with $n^C p_{1+}{}^T$.

This first step is not enough to identify the missing internal cells, however. A good way to see this is to use the framework developed by Shapiro (1982). The odds ratio for controls is

$$(4)\quad S^c_{\cdot} = \frac{m^c_{\cdot 1}}{m^c_{\cdot 0}} = \frac{m^c_{11}+m^c_{01}}{m^c_{10}+m^c_{00}} = \frac{m^c_{11}}{m^c_{10}+m^c_{00}} + \frac{m^c_{11}}{m^c_{10}+m^c_{00}}$$

$$= \frac{m^c_{10}}{m^c_{10}+m^c_{00}}\,\frac{m^c_{11}}{m^c_{10}} + \frac{m^c_{00}}{m^c_{10}+m^c_{00}}\,\frac{m^c_{01}}{m^c_{00}}$$

$$= \frac{m^c_{10}}{m^c_{10}+m^c_{00}}S^c_1 + \frac{m^c_{00}}{m^c_{10}+m^c_{00}}S^c_0.$$

When the relative frequencies of performance observed for the treatment group are substituted for the unobserved relative frequencies in (4), the relationship between the unknown performance-specific odds ratios $S_0{}^C$ and $S_1{}^C$ becomes

$$(5)\quad \frac{x^c_{\cdot 1}}{x^c_{\cdot 0}} = \frac{x^T_{10}}{x^T_{10}+x^T_{00}}\,\frac{m^c_{11}}{m^c_{10}} + \frac{x^T_{00}}{x^T_{10}+x^T_{00}}\,\frac{m^c_{01}}{m^c_{00}}$$

$$= \frac{x^T_{10}}{x^T_{10}+x^T_{00}}S^c_1 + \frac{x^T_{00}}{x^T_{10}+x^T_{00}}S^c_0.$$

In a graph of S_0 against S_1, relationship (5) is a downward sloping line for controls, while S_0 and S_1 are known points for assignees to the treatment. Since the parameters (2) and (3) are ratios involving points on this line, they are not identified without further restrictions.

A set of restrictions which would permit the parameters to be estimated is that the interaction structure between outcome and performance be the same for controls as for treatment assignees; in other words, that there is no three-factor interaction. Taking the four relative frequencies in the body of the upper half of the table and transferring them to the bottom half of the table, and then using iterative proportional fitting to make the control cells fit the control margins, would be one way to fill in the blanks.

This procedure was carried out on the wage subsidy evaluation data, resulting in Table 3. The missing control row margins were replaced with treatment group margins 91 and 51. Starting with the observed treatment counts for the treatment group from the top half of the table, iterative proportional fitting to both sets of control margins yielded the fitted counts in the bottom half of Table 3.

Table 3
Observed and Imputed Counts, by Assignment, Performance, and Outcome
Control Performance Imputed by "Uncollapsing"

Assignment Group (G)	Performance Indicator (Q)	Outcome Group: Earnings Target Not Reached (Y=0)	Outcome Group: Target Reached(Y=1)	Row Margins
Treatment(G=T)	NotPlaced(0)	83 = 147(.565)	13=147(.088)	96
	Placed (Q=1)	34 = 147(.232)	17=147(.116)	51
	Column Margins	117 = 147(.796)	30=147(.204)	147[a]
Control(G=C)	NotPlaced(0)	90.4=147(.615)[b]	5.6=147(.04)[b]	96[c]
	Placed (Q=1)	42.6=147(.290)[b]	8.4=147(.06)[b]	51[c]
	Column Margins	133 = 147(.905)	14=147(.095)	147

SOURCE: Calculations using State of Maine AFDC and Unemployment Insurance earnings records for the Training Opportunities in the Private Sector Program evaluation sample.

NOTES: Due to rounding, there may be discrepancies in sums and differences of relative frequencies.

The target earnings level was $2080 per quarter during the seventh through eleventh quarters after random assignment. This amount is equivalent to the earnings from full-time work at four dollars an hour.

[a]The size of the full treatment group was 297. A subgroup of 147 was selected using random order nearest available Mahalanobis metric pair matching to control group values of 14 pre-assignment covariates. Relative frequencies for the full treatment group were quite similar to those reported here.

[b]Imputed using iterative proportional fitting of treatment group relative frequencies to both sets of control group marginals.

[c]Imputed using treatment group row marginals.

IV. Pair Matching

One of the drawbacks of an approach which models and predicts performance is that results may be sensitive to the functional form of the model and its error distribution. Matching is more of a nonparametric alternative.[4] Rubin (1973) described various forms of matching on a single criterion variable and provided a FORTRAN program to carry out his recommended method, nearest avalilable pair matching. This method can easily be extended to allow matching on a large set of criterion variables by first computing the Mahalanobis distance between each member of the two sets of cases to be matched. After sorting the smaller set of cases into random order, Rubin's program can be used to choose a mate from the second set of cases for each case in the first set. Performance characteristics observed for one mate can simply be ascribed to the other mate, whose performance characteristics were not observed.

This procedure was carried out on the wage subsidy evaluation data. Fourteen variables measured at the time of random assignment were used to calculate Mahalanobis distances between each of the 147 assigned to the control group and each of the 297 assigned to the program treatment. The 147 control cases were sorted into random order, and each was matched with the nearest available treatment group case. Ascribing the oberved job placement status of each treatment assignee to her mate in the control group led to the data in Table 4.

Table 4
Observed and Imputed Counts, by Assignment, Performance, and Outcome
Control Performance Imputed by Pair Matching

Assignment Group (G)	Performance Indicator (Q)	Outcome Group: Earnings Target Not Reached (Y=0)	Outcome Group: Target Reached(Y=1)	Row Margins
Treatment(G=T)	NotPlaced(0)	83 = 147(.565)	13=147(.088)	96
	Placed (Q=1)	34 = 147(.232)	17=147(.116)	51
	Column Margins	117 = 147(.796)	30=147(.204)	147[a]
Control(G=C)	NotPlaced(0)	89 = 147(.605)[b]	7=147(.048)[b]	96[b]
	Placed (Q=1)	44 = 147(.299)[b]	7=147(.048)[b]	51[b]
	Column Margins	133 = 147(.905)	14=147(.095)	147

SOURCE: Calculations using State of Maine AFDC and Unemployment Insurance earnings records for the Training Opportunities in the Private Sector Program evaluation sample.

NOTES: Due to rounding, there may be discrepancies in sums and differences of relative frequencies.

The target earnings level was $2080 per quarter during the seventh through eleventh quarters after random assignment. This amount is equivalent to the earnings from full-time work at four dollars an hour.

[a]The size of the full treatment group was 297. A subgroup of 147 was selected using random order nearest available Mahalanobis metric pair matching to control group values of 14 pre-assignment covariates. Relative frequencies for the full treatment group were quite similar to those reported here.

[b]Imputed using pair matching.

V. Effects of Placement and Non-Placement

Using the column margins of Table 4, the overall effect of assignment to the wage subsidy program on attainment of the target earnings level can be calculated as 0.204 - 0.095 = 10.9 percentage points. The matching-method estimate of the effect on those who were placed is (17/51) - (7/51) = 19.6 percentage points, while the effect on those not placed is estimated as (13/96) - (7/96) = 6.2 percentage points. Similarly, using the "uncollapsed" imputations of Table 3, the effect on those who were placed is (17/51) - (8.4/51) = 16.8 percentage points, while the effect on those not placed is (13/96) - (5.6/96) = 7.7 percentage points. The fairly close agreement between the two sets of estimates is reassuring, since the two methods are based on very different underlying assumptions. Loglinear analyses of both tables failed to reject the hypothesis that the impact on those not placed was zero, and rejected the hypothesis that impacts on the two performance categories were the same.[5] These results certainly are plausible, indicating that the effect of the program is greater for those who completed it than for those who did not, and suggesting there was no effect on those who did not complete the program.

VI. Summary and Future Work

This paper has presented two new approaches to disaggregating program impacts by performance category. One approach makes two assumptions about the structure of the performance-by-outcome subtable in order to "uncollapse" the overall group-by-outcome table. The other approach used a nearest available matching algorithm to find a control counterpart for each assignee to treatment. The observed performance of her counterpart in the treatment group was substituted for the unobserved potential performance of each control. In an empirical application, the two very different methods yielded quite similar results.

Provided that outcome measures are categorical, extension to larger numbers of performance categories (more than two rows) is immediate. Nothing in either method limits the number of performance categories to two. For example, the category "placed" could be expanded into "placed within 7, 30, 60, 90, or more than 90 days" of random assignment, and then the problem could be analyzed with 2 x 2 x 6 contingency tables. The pair matching method, but not the "uncollapsing" approach, also may be used to stratify impacts on continuous outcomes into several performance categories.

References

Auspos, Patricia, George Cave, and David Long. Maine: Final Report on the Training Opportunities in the Private Sector Program. New York: Manpower Demonstration Research Corporation, 1988.

Brown, Randall. "The Effects of Length of Time Spent in Supported Work on Program Impacts. Appendix C of Rebecca Maynard, The Impact of Supported Work on Young School Dropouts. New York: Mathematica Policy Research and Manpower Demonstration Research Corporation, 1980.

Cave, George. "Noncompliance Bias in Evaluation Research: Assignment, Participation, and Impacts." Unpublished manuscript, January, 1988.

Chen, Tar, and Stephen E. Fienberg. "The Analysis of Contingency Tables with Incompletely Classified Data." Biometrics 32, 1 (March 1976): 133-144.

Devijver, Pierre A. "An Overview of Asymptotic Properties of Nearest Neighbor Rules." In Pattern Recognition in Practice, edited by E.S. Gelsema and L.N. Kanal. New York: North-Holland, 1980.

Ducharme, Gilles R., and Yves LePage. "Testing Collapsibility in Contingency Tables." Journal of the Royal Statistical Society, Series B 48, 2 (1986): 197-205.

Rubin, Donald B. "Matching to Remove Bias in Observational Studies." Biometrics 29, 1 (March 1973): 159-183; Printer's correction 32, 1 (March 1976): 955.

Shapiro, Stanley H. "Collapsing Contingency Tables--A Geometric Approach." American Statistician 36, 1 (February 1982): 43-46.

Simpson, S.H. "The Interpretation of Interaction in Contingency Tables." Journal of the Royal Statistical Society, Series B 13 (1951): 238-241.

1. George Cave is Senior Research Associate at Manpower Demonstration Research Corporation. The ideas expressed here do not necessarily represent the views of MDRC.

2. Brown (1980) used a Tobit model to predict performance; Cave (1988) developed a ratio estimator for use when one performance category has zero impacts by assumption.

3. See Auspos, Cave, and Long (1988).

4. See Devijver (1980).

5. Chi-square tests at the 0.05 significance level were applied to loglinear model parameters.