INCOMPLETE DRAFT: STRICTLY PRELIMINARY - DO NOT QUOTE.

Reducing Attrition Bias using Targeted Refreshment Sampling and Matching.

Peter Dolton

(University of Newcastle-upon-Tyne, IoE University of London and CEP, LSE)

October 2001

Abstract

This paper examines the possibility of reducing attrition bias in panel data using targeted refreshment sampling and matching. The targeted refreshment sampling approach consists of collecting new data from the original sampling population from individuals who would never usually respond to surveys. Using the propensity score matching method in conjunction with refreshment sampling it is suggested that the dropouts from a panel can effectively be replaced. The procedure allows us to identify underlying joint distributions in the data. The method is illustrated using data from the Youth Cohort Surveys in the UK which suffer 45% attrition in the second wave. A comparison of the results of this method with other techniques for attrition modeling suggest that the technique could be an effective way to overcome a substantial part of the bias associated with attrition.

Address for correspondence:
Professor Peter Dolton
Department of Economics
University of Newcastle
Newcastle-upon-Tyne
NE1 7RU

The author gratefully acknowledges: financial support from the ESRC grant no L134251001 , research assistance from Jonathan Mounsey, data collection by Yvonne Balfour, Elaine Robson and Nick Meagher and comments from seminar participants in Newcastle.

Reducing Attrition Bias using Targeted Refreshment Sampling and Matching.

Peter Dolton

(University of Newcastle-upon-Tyne, IoE University of London and CEP, LSE)

October 2001

1. Introduction.

The importance of panel data has grown remarkably in recent years. One of the perennial problems with panel data is that drop out from second and subsequent waves can be substantial. This can cause substantial bias in the modeling of outcomes in later waves if the analysis is based only on the respondents who are ever present. One important survey of young people in the UK over the last 15 years is the Youth Cohort Survey. Drop-out rates from this survey have averaged 45% and there is concern that this drop-out is not random and may be more prevalent amongst the less able and disaffected. If this is so, it is likely that modeling the outcomes of these young people using only those who respond at each wave, could bias the conclusions considerably.

This paper examines the possibility of using matching via propensity scores with refreshment sampling to reduce the effect of attrition bias in panel data. The motivation for the paper comes from the severe attrition experienced in many panel studies. Often it is suspected that the attrition is non-ignorable in the sense that the kind of people who will dropout of the sample will be non-random in both observable and unobservable characteristics. Of most concern are situations when the econometrician wishes to model an outcome variable – like labour market state – and believes that the observation of this variable may not be independent of the process of dropping out of the panel.

The idea of using imputation or matching methods to overcome missing data is not new. There are many methods for dealing with missing data on particular variables and on missing data relating to completely missing observations. These methods have recently become popular by economists seeking to evaluate policy interventions in a non-experimental context. The use of matching methods to attempt to attenuate the bias effects of attrition in panels has been fairly limited. The method of nearest neighbour hot deck imputation was suggested by Little and Rubin (1987) for precisely this problem but the idea has so far not been pursued.[1] In contrast, the idea of using refreshment or substitution samples to replace missing observations is much older. The essential idea was suggested by Kish and Hess (1959) over 40 years ago. There have been many alternative procedures suggested by different authors. These developments are admirably summarised by Rubin and Zanutto (2001). Most of these contributions discuss the possibility of finding direct substitutes for dropouts via random sampling of new respondents. We suggest a new method which involves using a targeted refreshment sample and matching from the respondents.

The suggestion of the paper is that it is possible to find better matches for the people who dropout of the panel by the use of a targeted refreshment sample. Most specifically we suggest that a targeted refreshment sample consisting of the kind of people who do not normally respond to questionnaires may be necessary in order to effectively replace the dropouts and estimate models of outcomes in which the bias associated with attrition has been minimised.

The paper examines the alternative estimators that have been suggested to address the attrition issue. We use the framework and notation of Hirano, Imbens, Ridder and Rubins (2001) for convenience and compare our estimates using targeted refreshment sampling and imputation with matching to those suggested by HIRR based solely on random refreshment sampling and MCMC imputation. The research described in this paper is crucially different from HIRR in a number of important respects. Firstly our refreshment sample is not a random sample – it is targeted at those who would not respond. Secondly we use the techniques of propensity score matching to allow the non-respondents to be matched with either those from the targeted refreshment sample or imputed from the population of respondents based on the similarity of observable characteristics. A further difference is that since we collect an entirely new sample as our refreshment sample then we retrospectively observe all outcome measures we are

[1] See p65 in Little and Rubin (1987)

interested in with respect to the dropout replacements and not just those which took place after the sample date.

The method of allowing dropouts to be matched to either new observations from the refreshment sample or existing units who are respondents at the second wave of the survey gives us a way of distinguishing between drop outs who do not 'look like' respondents and those who do. This is then instructive of a possible typology of attrition in the sense that there is heterogeneity in the sample of those who dropout of a panel study. There has been evidence of this typology of attrition in the work of Dolton, Lindeboom and Ven den Berg (2001).

Like the HIRR paper we will use the case of a binary outcome variable without conditioning regressors to examine the issue of identification. The method of matching also allows us to estimate the average treatment effect of attrition. This effect is the difference in the labour market outcome in the sample of those responding compared to those who drop out. We will then compare the estimates for the outcome equation with conditioning regressors using all the different estimators that have been suggested.

Our basic findings are that targeted refreshment sampling with imputation to match for drop-outs proves to be a more satisfactory method to overcome the estimation bias associated with attrition than many that have previously been suggested.

2. Basic Notation and the Sampling Framework

To simplify what follows we adopt the notation of HIRR. In our data we consider a three period model. The data is originally collected at the end of period 1 and contains information relating to period 0. Let Y_{it} be a vector containing all outcome variables of interest for individual *i* at time *t*. In the first period we draw a random sample of size N_P from a fixed population – we call this the *panel*. For each observation in this sample we observe X_i , Z_i and Y_{i1}. For a subset of size N_{BP} of this sample, who do not drop out of the follow up survey, we observe a second period outcome variable Y_{i2}. We refer to this sample as the *balanced panel* (BP). The remaining $N_{IP} = N_P - N_{BP}$ units who have dropped out of the panel and their Y_{i2} are missing; this group are called the *incomplete* *panel* (IP). In addition to the panel data set we draw a new random sample from the original population, of size N_R, which we call the *refreshment sample*.

We may formulate the data generation process in the manner of HIRR. Each unit in the population is assigned a 2 valued sampling indicator. If assigned $A_i = 2$, unit *i* is part of the panel and is approached both periods. If assigned $A_i = 1$, the unit is designed to be part of the refreshment sample and will only be approached in the second period.

We assume that all units respond the first time they are approached. Not all units respond in the second period they are approached. Let W_i be an indicator denoting the willingness to respond at the second wave of the panel. Hence $W_i = 1$ represents those units who respond in the second period and $W_i = 0$, those units who do not. This indicator is only observed if the researcher attempts to get a response from *i*. We can use the missing variable indicator and the design variable to define a missing data indicator D_{i2}.[2] When $D_{i2} = 1$ we observe Y_{i2} and when $D_{i2} = 0$ we do not observe Y_{i2}.

As we have described it so far the data generation process is very similar to that in HIRR. However, in our empirical data, we actually have two additional features which add interest and a little complexity. Firstly we have exogenous data merged from other sources which relates to each unit but comes from their school or geographical location. The value of this extra data is that it will potentially allow us to compare the use of exclusion restrictions in a HW framework to identify the attrition process.

Our second added complexity relates to the outcome variables. Our outcome state is actually measured each month in the form of a labour market state. Although our data collection took place on a given month we find out information about all previous months. Likewise at the second sweep of the panel we find out information at all previous months back to the first sweep but including an overlap period Y_{i1}^*. This overlap month is missing for the incomplete panel but present in the balanced panel and the refreshment sample. This variable is interesting as it will allow us to compare the end of year 1 state viewed at two points in time.

The whole of the data generating process is summarised in Table1 below.

---

[2] In HIRR they have two such indicators as in their data they do not observe y1 for the refreshment sample. In our case we do and hence we only need one missing data indicator.

Table 1: Summary of Data Generating Process.

Sample	Missing Data Indicator	Design Variable	Individual Characteristics						
	D2	A	W	Y0	Y1	Y1*	Y2	Z	X
Balanced Panel	1	2	1	√	√	√	√	√	√
Incomplete Panel	0	2	0	√	√	-	-	√	√
Refreshment Sample	1	1	-	√	√	√	√	√	√

The most important concern in the sampling position described above is whether it is possible to recover the joint distribution of (Y_1, Y_2, X) or possibly the conditional distribution of (Y_1, Y_2) given X. Of central importance in this search for identification is the specification of the attrition probability:

$$P(W_i = 1 | Y_1, Y_2, X)$$

One approach is to assume that

$$P(W_i = 1 | Y_1, Y_2, X)$$
$$A_i \perp W_i, Y_{i2}, Y_{i1,} X_i$$

in which case one can write the joint distribution of (Y_1, Y_2, X) as:

$$f(Y_1, Y_2, X) = \frac{f(Y_1, Y_2, X | W = 1). \Pr(W = 1)}{\Pr(W = 1 | Y_1, Y_2, X)}$$

Therefore, specification of the attrition probability under the assumption of random refreshment sampling, is sufficient to identify the joint distribution. Without this assumption one can proceed in one of two ways: either use the conditional probability of attrition to weight the complete panel (see Hansen, Hurwitz and Madow (1996) or Hellerstein and Imbens (1999) or use it to impute missing data (see Rubin 1987,1996)).

3. Models for Attrition in Panel Data

This section summarizes the various estimation models which address the problem of attrition in panel data.

3.1 MISSING AT RANDOM

The first model assumes that Y_{i2} is missing at random (MAR) in the panel i.e. that

$$W_i \perp Y_{i2} \Big| Y_{i1,} X_i \qquad \text{(MAR)}$$

This model implies that the missing data process is ignorable in the sense of Rubin (1976) and Little and Rubin (1987).

A more specific version of this model is the special case in which

$$W_i \perp Y_{i1}, Y_{i2,} X_i \qquad \text{(MCAR)}$$

In this model, referred to as missing completely at random (MCAR). In this case no bias results from simply using only the respondents at both waves to perform the analysis.

3.2. THE HAUSMAN-WISE MODEL FOR ATTRITION.

The basic principle of the Hausman and Wise (1979) model is that the probability of attrition in the second period depends on contemporaneous outcomes but not on first period outcomes. Hence we may write this as:

$$W_i \perp Y_{i1} \Big| Y_{i2,} X_i \qquad \text{(HW)}$$

The very simplest form of this model can be written in the following form:

$$Y_{i2} = X_i \beta + \varepsilon_i$$
$$W_i^* = Z_i \delta + v_i$$

where $W_i = 1 \quad if \quad W_i^* \geq 0; \quad W_i = 0 \quad if \quad W_i^* < 0$

W_i is a binary indicator for response $(W_i = 1)$ or non-response $(W_i = 0)$ and W_i^* is its underlying unobserved latent determinant; Z_i contains Y_{i2} and most or all of the elements of X_i and possibly some additional exogenous variables, and ε_i and v_i are mean-zero unobservables.

In the sample of respondents,

$$\begin{aligned} &E(Y_{i2} | W_i = 1, X_i, Z_i) = X_i \beta + E(\varepsilon_i | v_i \geq -Z_i \delta) \\ &= X_i \beta + h(Z_i \delta) \end{aligned}$$

The exact form of the bias on the conditional expectation of Y_{i2} depends on the function *h* which will depend explicitly on the assumed joint distribution of ε_i and v_i and the constituent elements of Z_i. The identification of the model will depend explicitly on these modeling assumptions.

3.3. THE ADDITIVELY NON-IGNORABLE HIRR MODEL.

HIRR suggest the model in the presence of a refreshment sample that a new random sample from the population is used to obtain data on the relation between Y_{i2} and X_i. Then via the process of data augmentation or imputation (See Tanner and Wong (1990) or Gelman and Rubin (1992)) which may use MCMC methods a new refreshed dataset is created which is complete in the sense of having a full set of real or imputed values on Y_{i1} and Y_{i2}. Then they suggest estimating the following limited dependent model.

$$P(W_i = 1 | Y_{i1} = y_1, Y_{i2} = y_2) = g(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2) \qquad \text{(AN-HIRR)}$$

They show that given a refreshment sample, a method of imputation and a choice of g it is possible to estimate the parameters $\alpha_0, \alpha_1, \alpha_2$. from which the model can be identified.

4. The Matching Approach.

The matching approach using propensity score methods was first suggested by Rosenbaum and Rubin (1983). They developed this methodology to facilitate the comparison between treated and untreated observations in a non-experimental setting. Matching estimators try to re-establish the condition of the experiment when no such data is available by choosing a comparison group from all the non-treated such that the selected group is as similar as possible to the treatment group in terms of their observable characteristics. The method has been recently applied extensively to the evaluation of training programmes see Dehejia and Wahba (1998, 1999) and Heckman, Ichimura and Todd (1999). The method is now being more widely applied to any situation in which a comparison is sought between two groups where the assignment to the groups is non-experimental and non-random. Hence the methodology can be applied to the assessment of the effect of attrition. This section sets out how the matching method using propensity scores can be applied to the problem of assessing attrition bias.

Using the potential outcome notation of Rubin (1974), let $Y_{i2}(W_i = 1)$ represent the outcome when i responds to the follow up survey (i.e. does not drop out of the sample) and let $Y_{i2}(W_i = 0)$ represent the outcome when i drops out. The fundamental problem in the assessment of attrition bias is that there is missing data since only $Y_{i2}(W_i = 1)$ or $Y_{i2}(W_i = 0)$ can be observed and we cannot observe the counterfactual for each person. In this context we can therefore write the realised outcome for i as:

$$Y_{i2} = W_i Y_{i2}(W_i = 1) + (1 - W_i) Y_{i2}(W_i = 0)$$

In order to facilitate ease of terminology we will use the terms familiar from the evaluation literature. Hence we may write the 'Treatment Effect' for unit i as:

$$\tau = Y_{i2}(W_i = 1) - Y_{i2}(W_i = 0)$$

Considering the whole population we can write the 'Average Treatment Effect' (ATE) as:

$$ATE = E\left[Y_{i2}(W_i = 1) - Y_{i2}(W_i = 0)\right]$$

where the first term on the right hand side is only observed for the respondents and the second term is only observed for the dropouts. The expectation is taken over the whole sample. Hypothetically this expression would give the average outcome difference effect in the sample of responding rather than dropping out. In practice of course we do not observe each individual in both of the two states and so without further assumptions the expression is impossible to evaluate. The crucial assumption underlying the propensity score method is the conditional independence assumption[3].

Definition 1. Rosenbaum and Rubin (1983): CONDITIONAL INDEPENDENCE ASSUMPTION (CIA). *Assignment to attrition (treatment), W_i is conditionally independent given pretreatment variables X_i if:*

$$W_i \quad \perp \quad Y_{i2}(W_i = 1), Y_{i2}(W_i = 0) \;\Big|\; X_i$$

This is a very strong and restrictive assumption as it requires that assignment to attrition (treatment) is associated only with observable variables and therefore that all the relevant differences between the drop-outs and respondents are captured in these observable attributes, and that conditional on them, attrition can be taken as random.

The power and attraction of the CIA assumption is that it validates comparisons for units with the same value of the covariates:

[3] Note that his assumption is variously called the 'unconfoundedness' or 'ignorable treatment assignment' assumption. For simplicity we refer to it as conditional independence.

$$E\left[Y_{i2}(W_i=1)\middle|W_i=1,X_i\right]=E\left[Y_{i2}(W_i=1)\middle|W_i=0,X_i\right]=E\left[Y_{i2}(W_i=1)\middle|X_i\right]$$
$$E\left[Y_{i2}(W_i=0)\middle|W_i=0,X_i\right]=E\left[Y_{i2}(W_i=0)\middle|W_i=1,X_i\right]=E\left[Y_{i2}(W_i=0)\middle|X_i\right]$$

The implication of this assumption then is that if we can match up the treated and non-treated people on their covariates then a conditional comparison (given these X_i) is valid. This allows us to establish a condition under which an average treatment effect (ATE) is identified.

Given the CIA assumption the population ATE is identified:

$$\begin{aligned} ATE &= E\left[Y_{i2}(W_i=1)-Y_{i2}(W_i=0)\right] \\ &= E\left[E\left[Y_{i2}(W_i=1)\middle|W_i=1,X_i\right]-E\left[Y_{i2}(W_i=0)\middle|W_i=0,X_i\right]\right] \end{aligned}$$

where the outer expectation is taken over X_i.

In principle one could make this comparison operational if there were a small discrete number of possible covariates with a limited number of values. In this situation with a large number of observations in the dataset there would be a sufficient number of people in each cell to do the matching. When this is not the case the alternative of the propensity score is useful.

Definition 2. *The propensity score is the conditional probability of attrition given the period 1 exogenous variables:*

$$p(X_i)\equiv \Pr(W_i=1|X_i)$$

The propensity score has two important properties:

Lemma 1 THE BALANCING PROPERTY.

$$W_i \perp X_i \mid p(X_i)$$

This property asserts that attrition and the observed covariates are conditionally independent given the propensity score. Combined with the CIA assumption the balancing property suggests the key property of the propensity score:

Lemma 2 .CIA GIVEN THE PROPENSITY SCORE.
If assignment to attrition is conditionally independent then assignment to attrition is independent given the propensity score.

$$W_i \quad \perp \quad Y_{i2}(W_i = 1), Y_{i2}(W_i = 0) \;\Big|\; p(X_i)$$

R&R show that you can identify the ATE using the propensity score with matching. Hence the ATE can be written as:

$$ATE = E\Big[E\big[Y_{i2}(W_i = 1)\big|W_i = 1, p(X_i)\big] - E\big[Y_{i2}(W_i = 0)\big|W_i = 0, p(X_i)\big]\Big]$$

The implication of this is that matching on the propensity score can be used to replace those who drop out of the sample with those who 'look like' them but remain in the sample. Obviously it should be stressed again that the validity of this expression is crucially dependent on the restrictive CIA assumption.

5. Targeted Refreshment Sampling and Matching.

The idea of targeted refreshment sampling is that we obtain a totally new sample of people from the population who were not part of the original panel and who would never normally respond to a survey or who would drop out of a survey if contacted initially. The purpose of obtaining a targeted sample of these people is that we can selectively replace those 'hard to match' individuals in the incomplete panel who are difficult to match. Hence the distinctive feature of our procedure is to allow the dropouts to be matched not only to those in the balanced panel but also to those in the targeted refreshment sample who look like them.

Definition 3. TARGETED REFRESHMENT ON NON-RESPONDENTS.

Define $W_j^* = \begin{cases} W_i & \forall \quad i \in BP \\ 0 & \forall \quad j \in M, R \end{cases}$

Where M is the set of drop out matches from BP. Hence W^* is the same as W except that for those from the refreshment sample and those matched from the balanced panel to substitute for the incomplete panel are assumed to be non-respondents.

Now if we define the CIA assumption for this targeted refreshment sample:

Definition 4. TARGETED REFRESHMENT CIA.

$$W_j^* \perp Y_{j2}(W_j^* = 1), Y_{j2}(W_j^* = 0) \mid p(X_j)$$

This assumption requires the modified attrition status (according to the targeted refreshment on non-respondents) is conditionally independent of the outcome given the propensity score.

Then it is a straightforward corollary of R&R to show that the ATE can be evaluated by using the propensity score matching technique to replace dropouts either from the refreshment sample or the balanced panel;. Hence we can rewrite the ATE in the targeted refreshment sample as:

$$\begin{aligned} ATETRSM &= E\left[Y_{i2}(W_i = 1) - Y_{i2}(W_i = 0)\right] \\ &= E\left[E\left[Y_{j2}(W_j^* = 1) \mid W_j^* = 1, p(X_i)\right] - E\left[Y_{j2}(W_j^* = 0) \mid W_j^* = 0, p(X_j)\right]\right] \end{aligned}$$

What this expression says is that the ATE of attrition can be evaluated using the Targeted Refreshment Sample and matching.

Several results are clear from this procedure:

1. The match quality from using the TRSM approach can never be worse than that from the BP alone as, by definition we are allowing the algorithm to match to any of the original BP or the R sample. Hence an R sample replacement will only be chosen if it is 'closer' to the dropout than anyone from the BP.
2. The more accurately the refreshment sample approximates the IP then the larger the fraction of the matching which will come from this R sample.
3. The part of the R sample which is used in the matching process is unlike any person in the BP sample and hence is instructive of the diversity in the drop out population.
4. We would expect that the TRSM approach to give parameter estimates which are between the extremes of the MCAR and HW model as the assumptions involved are not so extreme as suggesting that attrition is at random or completely determined by contemporaneous events.

So far we have not specified how the matching procedure algorithm is to be used. In practise there are many different matching algorithms[4]. Since these computational details are not the subject of this paper we will focus on the simplest method.[5] We will therefore focus on the using nearest neighbour matching with replacement partly because this has the attractive feature of having one real matched individual for each dropout from the panel. This is most convenient since we actually want to compare the BP population with that generated by the TRSM technique. This means we seek to replace each dropout with one control from the BP or R sample. In this case what we seek is the closest comparable person in terms of propensity score. In formal terms we seek to match each drop out unit *i* with someone from the BP or R sample such that:

[4] See Rosenbaum and Rubin (1985), Lechner (2000) and Todd (2000) for practical details of different matching methods.

[5] Using kernal and local linear regression methods with our data do not change the conclusions.

$$p(X_i) - p(X_j) = \min_{k \in \{W_j = 0\}} \left\{ \left| p(X_i) - p(X_k) \right| \right\}$$

Clearly if we use this nearest neighbour match there is the question of what constitutes a close enough match. This amounts to how tight the caliper bounds need to be set on the nearest neighbour matching. What is appropriate in this context is unclear. We use the percent reduction in bias for caliper matching tables in Cochran and Rubin (1973) based on the ratio of variances of Y_{i2} in Table 2.3.1. Using this table we use a caliper of .03 for all our matching computations.

6. The Simple Binary Model

The analysis of the attrition model can be specified for the specific case where the outcome variables Y_{i1} and Y_{i2} are binary and we suppress the conditioning on time invariant co-variates X_i. Denote the conditional probability

$$q_{yw} = P(Y_{i2} = 1 | Y_{i1} = y, W_i = w)$$

and the probability

$$r_{yw} = P(Y_{i1} = y, W_i = w)$$

In large samples we can learn the value of r_{yw} for $y, w \in \{0,1\}$ since the original sample is a random sample from the population.

The restrictions on the estimable attrition models are specified in HIRR. They show that:

MAR: $q_{00} = q_{01}$ *and* $q_{10} = q_{11}$

HW: $$\hat{q}_{00} = \frac{r_{10}.r_{01}.(1-q_{01}) - r_{11}.r_{00}(1-q_{11})}{r_{00}.r_{11}.q_{11}.(1-q_{01})/q_{01} - r_{11}.r_{00}.(1-q_{11})}$$

And

$$\hat{q}_{10} = \frac{q_{00}.r_{00}.q_{11}.r_{11}}{q_{01}.r_{01}.r_{10}}$$

From Theorem 1 in HIRR (1998) we can write:

AN: $$\hat{q}_{00} = \left\lfloor q_{01}r_{01}/g(\boldsymbol{a}_0 + \boldsymbol{a}_2) - q_{01}r_{01} \right\rfloor / r_{00}$$

$$\hat{q}_{10} = \left[\frac{r_{10} + (1-q_{11})r_{11} - {(1-q_{11})r_{11}}/{g(\boldsymbol{a}_0 + \boldsymbol{a}_1)}}{r_{10}} \right]$$

From direct estimation of the AN model using our data augmented by using the TRSM method we estimated $a_0 = .457, a_1 = -.354, a_2 = -.463, a_0 + a_1 = .103$. These results allow us to retrieve the AN estimates of $\hat{q}_{00}$ and $\hat{q}_{01}$

If we define $t \in T$ as the set of all matched units from BP and R i.e. the TRSM sample associated with W_t^* then the TRSM estimates of $\hat{q}_{00}$ and $\hat{q}_{01}$ can be estimated directly from the data after the TRSM procedure from:

$$q_{yw^*} = P(Y_{i2} = 1 | Y_{i1} = y, W_t^* = w)$$

After we have introduced the data we will use these estimators to directly compare the alternative methods of identifying the underlying structure of the joint distribution of the outcome variables and attrition identifier.

7. YCS Data and the Refreshment Sample.

The empirical application of the refreshment sampling ideas we are concerned with is to address the problems of the attrition in the Youth Cohort Surveys in the UK. We examine the YCS 7, 8, and 9 and focus on the people in these data who live in the North East of England. The reason for doing this is to make the collection of the refreshment sample feasible. The are 1829 young people in the YCS cohorts 7, 8 and 9 in the North East of England. The attrition rate from this sample at the end of the first wave of the sample is 45%. This leaves a Balanced Panel of 1008 for this dataset. Consequently there are 821 people in the Incomplete Panel.

Our refreshment sample consists of 880 people. These individuals were traced in the North East of England in a number of ways. We contacted: hostels for young people, sheltered accommodation, people on special programmes, targeted leavers from schools with no known destination. In addition we sampled young people on various government schemes for the young unemployed and disaffected. Most importantly, we also collected data on people in the New Deal for Young People. A sample of 502 young people were interviewed on the Routes project. These interviews averaged 45 minutes in length and retrieved all of the same information as the YCS postal questionnaire. The interviews were mainly conducted by tracing their whereabouts and interviewing them in their hostel or accommodation or government training establishment. A second major source for the refreshment sample was young people on the New Deal for Young People scheme in the North East. We collected data on another 378 from this source. These young people were asked to fill in a questionnaire in groups of 2-10 individuals. Whilst they were completing the questionnaire there were always one or two researchers there to help them record the information. In many cases the respondent could not read and the researcher had to conduct what amounted to a face to face interview. On average these questionnaires took around 15-20 minutes to complete. Most of the respondents completed the questionnaire in Newcastle College although those on the employment option were traced and interviewed in their place of work. Again the questionnaires contained all the same information as that in the YCS surveys.

The data in the YCS surveys is relatively rich. It includes information on : basic personal characteristics like gender and ethnicity, home family characteristics like parents education and occupation, numbers of brothers and sisters and where the young person lives, individual's school performance in terms of English and Maths GCSE, overall GCSE point score and truancy record, individual's labour market anticipation in terms of whether the person had a careers interview and whether they had any labour market experience whilst at school. The dependent variable of concern to us is whether the person is in a job or full time education at a specific point in time. Clearly we observe this for all the panel at the beginning of the data collection. But we do not observe this for those who drop out at the beginning of the wave 2 follow-up.

In addition to the sample data collected directly on the questionnaire (or at interview in the case of the Routes data) we devoted a lot of effort to the collection of exogenous data relating to these young people. Firstly since the LEASIS data contains the school DfES number of the school attended for each person we were able to use this to link the data with the LEASIS data supplied by the DfES. This data contains details of average class size, the proportion of children in the school who have special educational needs, the proportion excluded from the school for bad behaviour and the average GCSE score of the pupils in the school. We also linked our data to the data on the LEA in which the child is educated to retrieve the average GCSE points score of that local education authority. Finally since we have the postcode of the school attended we can also link the information on the local labour market. Most specifically we attached the dperivation scores and the level of local unemployment to the area of each child.

All the details of the samples, the definitions of the variables and their summary statistics are reported in Table A1 and Table A2 respectively. Looking at the summary statistics in Table A2 we see that there are major differences in the different subsamples. We return to this issue later when we report on concordance measures of mean covariate imbalance in the different samples.

It should be stressed that due to the different data collection methods from the YCS to the Routes and NDLP data we may expect some inaccuracies in the responses. For example responding to a question asking to self report truancy behaviour is a different experience if it is a written response on a postal questionnaire to a verbal response to a question posed in a one-to-one interview. In the former the individual may be inclined to be more honest but in the latter the person may be intimidated by the 'owning up' element to reporting their truancy to another person face-to-face. These factors must be borne in mind in any consideration of the limitations of the study.

Further caution should be expressed about the measurement of the outcome variable in this data. For the refreshment sample we measure this outcome retrospectively by asking individuals to remember what they were doing one year and 2 years after the statutory school leaving age of 16. In the YCS postal questionnaire individuals are asked to recall their labour market state each month over the previous year. It may well be that individuals will find this information, however asked, difficult to recall. It is of course possible that the degree of accuracy with which they recall this information is not independent of whether the questions come in the form of a paper questionnaire or a face-to-face interview. However, we were left with little choice of how we gathered information from the refreshment sample as they would not have replied to a postal questionnaire. In addition, we have no *a priori* beliefs about the direction of any bias in reporting which may result from the different methods of data collection.

Table 2 shows exactly the data in our balanced panel, incomplete panel and raw refreshment sample are configured. The table shows the joint distribution of Y_{i1} and Y_{i2}.

Table 3 shows the same information for the matched refreshment sample and also gives the breakdown of the Y_{i1} and Y_{i2} variables with the attrition indicator W_j.

Table 4 reports on the basic attrition equation which is used to compute the propensity scores. Some consideration of these results is worthwhile as it helps us to characterise the attrition problem and understand the matching process. Table 4 reports two separate equations. The first includes all the observations but not all the regressors as information on some of the exogenous data is lost in the merging process. The second reports on the attrition equation when all the variables are used but some of the observations are lost. The tables show us that individuals are more likely to drop out of the sample if: they are male and white, live away from home and their parents, attended a school with higher proportions of children excluded, and if they played truant more often from school. They are less likely to dropout if they have a higher GCSE score. They are also curiously more likely to drop out if their school's GCSE grades were higher or the GCSE grades in the LEA were higher. It is possible that one could interpret this as a relative effect, i.e. if the grades in the school or LEA are higher then one's own grades are likely to be seen as relatively lower. This may serve alienate people who are lower achievers and induce them not to respond to questionnaire follow ups.

Two important questions are exactly how different are the balanced panel and the incomplete panel and exactly how closely does the refreshment sampling with matching do in terms of matching up the original sample and replacing the IP. There is no clearly defined procedure for answering such questions. In Table 5 we adopt the informal concordance method of making comparisons suggested by Rosenbaum and Rubin (1985). This consists of tabluating the standardized difference in percent which is the mean difference as a percentage of the average standard deviation, ie.:

$$\left[\frac{100(\bar{x}_1 - \bar{x}_c)}{(s_1^2 + s_c^2)/2}\right]^{1/2}$$

where for each covariate $\bar{x}_1$ and $\bar{x}_c$ are the sample means in the first group and the 'control' or comparison groups and s_1^2 and s_c^2 are the corresponding sample variances.

We present this measure of concordance for each variable and each interesting comparable data set. Most importantly we can see from this table that the BP and IP are different by using this standardised concordance measure by around 21%. In contract if we compare the BP& IP with the BP&R/I targeted refreshment sample we see that this standardised difference comes down to an average of 6%. This informal measure suggests that the distance between the data sets has narrowed considerably by using the targeted refreshment sample approach. Note however that the one could not use the targeted refreshment sample alone to replace the IP as the difference between the BP and R is around 33%.

Table 2: Summary Statistics for YCS and Raw Refreshment Sample.

Sample	Y0	Y2	W	No of obs
Balanced Panel	0	0	1	108
N_{BP} =1008	0	1	1	99
	1	0	1	75
	1	1	1	726
Incomplete	0	-	0	263
Panel N_{IP} =821	1	-	0	558
Refreshment	0	0	-	499
Sample	0	1	-	80
N_{R} =880	1	0	-	140
	1	1	-	161

Table 3: Summary Statistics for YCS and Matched Refreshment and Imputation Sample.

Sample	Y0	Y2	W	No of obs
Balanced Panel N_{BP} =1008	0	0	1	108
	0	1	1	99
	1	0	1	75
	1	1	1	726
Incomplete Panel N_{IP} =821	0	-	0	263
	1	-	0	558
Matched only from Refreshment Sample N=306	0	0	-	165
	0	1	-	36
	1	0	-	47
	1	1	-	58
Matched from BP Sample N=515	0	0	-	67
	0	1	-	58
	1	0	-	37
	1	1	-	353
Matched from Refreshment and BP Sample N_R =821	0	0	-	232
	0	1	-	94
	1	0	-	84
	1	1	-	411

8. Simple Binary Model: Identification and Recovering Joint Distributions.

Using the different assumptions about the attrition process the estimators for the recovery of the joint distribution of the outcome and attrition indicator were set out in section 6 above. In this section we use the our data to compare these different estimators. The interest in this comparison is that it makes clear the necessary restriction on identification which is being assumed in order to recover the underlying distribution on the variables of interest.

Denote the conditional probability

$$q_{yw} = P(Y_{i2} = 1 | Y_{i1} = y, W_i = w)$$

and the probability

$$r_{yw} = P(Y_{i1} = y, W_i = w)$$

section 6 explained that $\hat{r}_{00}$, $\hat{r}_{01}$, $\hat{r}_{10}$, $\hat{r}_{11}$, $\hat{q}_{01}$, $\hat{q}_{11}$ were recoverable directly from the raw data. We tabulate these values for our data in Table 6.

Table 6. Probabilities $\hat{r}_{00}$, $\hat{r}_{01}$, $\hat{r}_{10}$, $\hat{r}_{11}$, $\hat{q}_{01}$, $\hat{q}_{11}$ Given by DGP:

	w=0	**w=1**	**qy1**
Y1=0	0.144	0.113	0.478
Y1=1	0.305	0.438	0.906

The probabilities $\hat{q}_{00}$, $\hat{q}_{10}$ not given by DGP that need modeling assumptions for identification are estimated and reported in Table 7.

Table 7. Probabilities $\hat{q}_{00}$, $\hat{q}_{10}$ Not Given by DGP.

	MAR	**HW**	**Raw R**	**Matched from R**	**Matched from BP**	**HIRR AN**	**Matched R&BP**
q_{00}	0.478	0.214	0.138	0.161	0.420	0.373	0.288
q_{10}	0.906	0.743	0.535	0.733	0.903	0.886	0.830
$q_{10} - q_{00}$	0.428	0.529	0.397	0.572	0.483	0.513	0.542

The first row in this table gives the probability of being in a good state in period 2 given non-response and a bad state in period 1. The second row of the table gives the probability of being in a good state in period 2 given non-response and a good state in period 1. The final row $q_{10} - q_{00}$ gives the difference in the probability (conditional on dropout) of getting into a good state at period 2 given a poor state at period 1 and a good state at period 1.

Not surprisingly we can see that there are substantial differences in these probabilities given the different estimation methods. We can see that the MAR and Matched from the BP estimates are very high and close to each other as they both assume that the BP reflects the IP in terms of outcomes. We know this to be an naïve assumption and hence a gross overestimate of the probabilities involved. Likewise basing our estimation solely on the targeted refreshment sample on the assumption that all dropouts are disaffected deadbeats is also too simplistic. Again it matters little if we take the refreshment sample per se or match from it – the estimates of the probabilities will be underestimates.

Table 8. Average Treatment Effect

	Raw R	Matched from R	Matched from BP	Matched R&BP
ATE	0.544	0.358	0.039	0.203

The ATE reported in Table 7 is the average outcome difference effect in responding rather than dropping out of the sample. As one would expect, if the nearest neighbour method of matching is forced to choose from the BP sample for a replacement, the attrition effect on Y_{i2} is only around 4%. In contrast, allowing matching from the R sample given a very targeted refreshment is likely to overstate the attrition effect as we can see by the 36% estimate in Table 7. The estimate of the ATE using the matched R &BP sample is 20%. This is much more plausible given the basic difference in the Y_{i2} variable between the BP and IP is around 11%.[10]

[10] See Table A2.

9. Effects of Estimation in More General Models.

One of the important advantages of the TRSM technique is that it permits us to replace all the dropouts in the original sample and then estimate a model of the outcome equation of interest. We report these results in Table 9 for different matched samples.

We focus here on the difference between the estimates one obtains when the outcome model is estimated using only the BP with those obtained when the BP is augmented using the TRSM method. In this comparison we can see that some coefficients on regressors become significant when previously they were insignificant and a group of others change the size of the coefficients markedly. In the former group are the variables: Gender, Ethnic, Living with Parents, % Special Needs. These results suggest that if only the Balanced Panel was used to estimate the relationship between the outcome variable and its explanatory factors then we would mistakenly believe that a person's gender and ethnicity was unimportant in the determination of the labour market state. In contrast we know from the TRSM procedure that males and those from the ethnic minorities are more likely to be in a good state at 18. Likewise those young people living at home with their parents are more likely to be in a good state as are those who attended a school with a lower proportion of children with special educational needs. All of these results would not come to light by using only the balanced sample to estimate the period 2 labour market equation.

The second category of effects of using the TRSM method is that it suggests that certain coefficients are misrepresented in terms of their importance. The table shows us that truancy, and the overall GCSE score are systematically overestimated in terms of their importance if one uses only the BP for estimation. In contrast the effect of having a Maths GCSE is grossly underestimated.

All of the effects of using the TRSM method are important and potentially unpredictable *a priori*.

10. Conclusions.

It is well known that attrition causes bias in outcome modeling in incomplete panels. This panel has reviewed the different approaches to this problem. The different methods of overcoming this problem were compared. Each of the methods can be seen as making some simplifying assumption in order to try to solve the identification problem posed by the individuals who drop out of the sample.

We showed that the MAR and MCAR and HW methods often seem inappropriate models – estimates from these models are wide apart and show that the nature of the original assumption is too naive with respect to the nature of the incomplete panel.

The HIRR AN model was shown to be more appropriate in the sense of giving estimates which are between the extremes of the MCAR and HW models. Our suggested TRSM model is a logical alternative to the AN model and is shown to have similar estimates. But the matching component shows that using only targeted refreshment sample could be wrong.

The TRSM model offers a number of advantages over the AN model. Firstly its use is suggestive of the nature of the incomplete panel. The higher the proportion of the refreshment sample which is used in the matching process compared to those matched from the balanced panel the more heterogeneous is the dropout population. Second, the use of the TRSM approach is suggestive of the typology of attrition. Specifically matching from both the balanced panel and the refreshment sample implies that drop-outs are heterogeneous – the characterisation of drop-outs as ‘deadbeats’ and ‘movers’ may be useful. A third advantage of the TRSM approach is that it can be used with the simple collection of additional data. How good the refreshment sample is will be determined completely objectively by the matching process. If the refreshment sample is not informative and does not match up with the drop outs at all then all the matches will come from the balanced panel to replace those who drop out. If, on the other hand, a substantial fraction of those replaced come from the refreshment sample rather than the balanced panel then we know we are characterising the drop out individuals better.

Further advantages of the procedure are that it provides easily comparable estimates to the other methods of correcting for attrition in the sense that we can compare the results of this procedure with any other alternative. In addition, since it replaces the dropout people it allows us to continue the modeling of the outcome variable with a 'full set' of 'bias corrected' data and hence permits the estimation of fully specified equations to explain subsequent labour market outcomes.

One implication of the diverse typology of people who drop out of panels is that it is not surprising that instruments are difficult to find to identify HW model. So if we see that attrition is caused not simply by one, easy to characterise, type of person – say a 'deadbeat' – but is rather common among both 'deadbeats' and 'movers' then it is not surprising that it is difficult to find instruments that correlate with dropping out, but do not correlate with the subsequent outcome variable.

The higher is the proportion of the matched sample which comes from the refreshment sample the more useful and important is the refreshment sample and the more biased will be any procedure which matches only from the balanced panel or uses structural assumptions to try to overcome the attrition problem.

The important implication of this paper for data collection agencies and survey organisations is that it would be prudent to collect new targeted refreshment samples to attempt to counteract the bias that results from attrition. Government agencies would do well to budget for the collection of extra refreshment samples to correct for attrition bias in panel data. There is considerable evidence of the bias that may be caused by attrition and this paper shows that targeted refreshment samples may attenuate the bias. This is most effectively achieved by focussing new data collection on the kind of individuals who would never normally respond to questionnaires. It appears that many of the dropouts from a survey will 'look like' respondents (in a matching sense) and can be matched or replaced by matching methods. Harder to find are substitutes for the kind of individuals the hard core of dropouts. It is suggested that survey research teams need to look hard to find these types of people to correct for attrition bias.

Table 4: Attrition Equation Estimates.

Pr(Dropout W=1)	Coef.	Std.	Coef.	Std.
Basic Personal Characteristics				
Gender	**-0.2083**	0.0616	-0.2149	0.2186
Ethnic	**-0.5075**	0.2138	**-0.7214**	0.2995
Home/Family Characteristics				
Fathqual	-0.0176	0.0812	-0.0319	0.0901
Mothqual	-0.0616	0.0838	-0.1202	0.0948
Mothjob	-0.0182	0.0712	-0.0408	0.0784
Fathjob	0.0481	0.0796	0.0786	0.0883
Siblings	0.0467	0.0294	0.0543	0.0334
Home	***-0.1499***	0.0774	***-0.1589***	0.0852
Livepare	-0.3434	0.1809	**-0.3974**	0.2031
Individual's School Performance				
English	-0.1220	0.0896	-0.1345	0.0978
Maths	-0.1184	0.0859	-0.1368	0.0947
Gcse	**-0.0060**	0.0029	***-0.0056***	0.0033
Truant	**0.2243**	0.0644	**0.2079**	0.0705
Individual's Labour Market Anticipation				
Careers	**-0.2369**	0.0946	**-0.2344**	0.1084
Workexp	0.0804	0.1055	-0.0755	0.1320
School Characteristics				
Avcl1415			0.0025	0.0158
Specwpct			**0.0659**	0.0319
Permexcl			**40.8588**	11.2446
Average_			**0.0143**	0.0064
LEA Exam Performance				
Leaavpts			**0.0852**	0.0155
Locality Characteristics				
Depind			-0.0035	0.0029
Unemrate			-0.0062	0.0226
Constant	**0.7557**	0.2070	**-2.8057**	0.6517
Nobs	1829		1566	
Chi	131.15		190.17	
Log likelihood	-1192.61		-985.625	
Pseudo R2	0.0521		0.088	

Table 5: Standardised Percentage Mean Covariate Imbalance in Different Matched Samples

Variable	BPvIP	BPvR	BPvR/I	IPvR/I	(BP&IP)v(BP)	(BP&IP)v(BP&R/I)
Labour Market State						
Y0	**26.33**	**59.85**	**42.71**	**16.04**	**-12.26**	**7.72**
Y1	**28.21**		**69.80**		**-13.14**	**18.94**
Basic Personal Characteristics						
Gender	**16.63**	-2.31	**9.30**	-7.31	-7.48	-3.29
Ethnic	**10.85**	**-52.74**	3.13	-7.81	-4.46	-3.09
Home/Family Characteristics						
Fathqual	**12.75**	5.48	**15.93**	3.18	-5.58	1.34
Mothqual	**12.98**	3.58	**10.07**	-2.92	-5.67	-1.24
Pared	**12.21**	3.59	**14.69**	2.47	-5.40	1.07
Mothjob	**9.77**	**36.39**	**17.37**	*7.56*	-4.40	3.42
Fathjob	**11.64**	**61.41**	**22.13**	**10.44**	-5.30	4.83
Siblings	-5.16	**-60.18**	**-24.16**	**-19.02**	2.37	**-9.37**
Home	**26.18**	**57.10**	**37.76**	**11.37**	**-12.13**	5.42
Livepare	**15.26**	**53.69**	**37.59**	**23.81**	**-7.71**	**13.60**
Individual's School Performance						
English	**36.96**	**32.90**	**32.69**	-4.15	**-16.84**	-1.92
Maths	**35.11**	**48.39**	**29.72**	*-5.25*	**-15.53**	-2.30
Gcse	**41.92**	**66.02**	**44.83**	4.06	**-18.85**	1.85
Gcsediff	**-40.06**	**-44.43**	**-27.63**	**12.38**	**18.24**	6.16
Truant	**-27.08**	2.16	-0.90	**26.17**	**12.32**	**11.91**
Individual's Labour Market Anticipation						
Careers	**14.70**	**49.05**	**22.57**	7.91	-6.92	3.92
Workexp	0.07	**22.18**	**14.66**	**14.59**	-0.03	**6.87**
School Characteristics						
Avcl1415	**-13.68**	**14.94**	1.38	**15.49**	6.16	6.82
Specwpct	**-13.31**	**-33.40**	**-20.66**	**-15.97**	**8.16**	**-10.22**
Permexcl	**-21.93**	**-48.62**	**-16.97**	4.52	**10.26**	2.49
Average_	4.71	**44.84**	**14.25**	**10.82**	-2.05	4.55
LEA Exam Performance						
Leaavpts	**-44.28**	**35.08**	**13.94**	**57.05**	**19.53**	**25.40**
Locality Characteristics						
Depind	0.47	**-15.53**	-8.83	-9.25	-0.21	-4.03
Unemrate	**-13.82**	**15.00**	-1.89	**12.15**	6.29	5.53
Predicted Attrition Probability						
Score	**-79.70**	**-68.73**	**-67.16**	**22.65**	**35.26**	-0.57
Average Standardized Covariate Imbalance	21.32	33.42	18.68	11.50	8.74	6.43
% of Significantly Different Covariates	80.00	80.00	80.00	60.00	56.00	32.00

Bold denotes significantly different on a t- test at the 5% level.

Table 9: Labour Market Outcome Equation Results with Different Samples.

	Balanced Panel				**Balanced Panel & Refreshment Sample**				**Balanced Panel & Refreshment and Imputation Sample**			
Y2	**Coef.**	**Std.**	**Coef.**	**Std.**	**Coef.**	**Std.**	**Coef.**	**Std.**	**Coef.**	**Std.**	**Coef.**	**Std.**
Basic Personal Characteristics												
Gender	**0.3342**	0.1073	0.5255	0.4039	**0.2314**	0.0715	**0.9276**	0.3598	**0.1508**	0.0749	0.1039	0.3281
Ethnic	0.3921	0.3522	0.2443	0.3776	**0.6045**	0.1312	**1.0274**	0.1837	**0.5212**	0.2441	***0.5301***	0.3101
Home/Family Characteristics												
Pared	-0.0066	0.1238	0.0445	0.1345	**0.2635**	0.0776	**0.2521**	0.0937	0.0851	0.0870	0.0738	0.0992
Mothjob	0.1642	0.1303	0.1566	0.1424	0.0005	0.0797	0.0216	0.0951	0.0777	0.0850	**0.2028**	0.0973
Fathjob	0.1129	0.1424	0.1099	0.1546	**0.1926**	0.0792	0.1029	0.0953	0.0814	0.0915	-0.0038	0.1059
Siblings	0.0312	0.0547	0.0246	0.0593	**-0.0737**	0.0258	**-0.1638**	0.0313	-0.0383	0.0300	-0.0465	0.0351
Home	-0.0174	0.1309	-0.0176	0.1419	0.2293	0.0917	0.1454	0.1063	0.1432	0.0922	0.0598	0.1034
Homesamp					-0.0556	0.1682	**-0.4420**	0.2201	0.0782	0.2843	-0.3201	0.3704
Livepare	0.0505	0.3401	-0.0718	0.3997	**0.4761**	0.1350	**0.4174**	0.1702	**0.6290**	0.1482	**0.6756**	0.1871
Livsamp					-0.2353	0.1518	***-0.3425***	0.1971	**-0.8226**	0.2155	**-0.7232**	0.2599
Individual's School Performance												
English	**0.4106**	0.1409	**0.4086**	0.1538	**0.3493**	0.0959	**0.2832**	0.1105	**0.4208**	0.1000	**0.3857**	0.1113
Maths	**0.3699**	0.1485	**0.3748**	0.1622	**0.3243**	0.0926	**0.2569**	0.1108	**0.4349**	0.1037	**0.3946**	0.1157
Gcse	**0.0215**	0.0053	**0.0186**	0.0058	**0.0141**	0.0031	**0.0173**	0.0037	**0.0173**	0.0035	**0.0160**	0.0040
Truant	***-0.2067***	0.1085	**-0.2418**	0.1167	**-0.1868**	0.0762	**-0.2301**	0.0868	**-0.1886**	0.0771	**-0.2509**	0.0853
Individual's Labour Market Anticipation												
Careers	-0.1259	0.1835	-0.0374	0.2029	**0.3088**	0.0980	0.1075	0.1293	0.1028	0.1137	***0.2380***	0.1332
Workexp	-0.0552	0.1928	0.0368	0.2276	***0.2056***	0.1089	0.2142	0.1456	0.0509	0.1212	0.1514	0.1513
School Characteristics												
Avcl1415			0.0303	0.0255			0.0150	0.0203			0.0232	0.0193
Specwpct			-0.0529	0.0520			-0.0303	0.0433			**-0.0792**	0.0389
Permexcl			**-53.7398**	19.1804			**-57.0766**	13.7176			**-57.4863**	13.7573
Average_			-0.0015	0.0105			-0.0021	0.0071			0.0092	0.0073
LEA Exam Performance												
Leaavpts			-0.0222	0.0256			0.0065	0.0179			0.0198	0.0177
Locality Characteristics												
Depind			0.0034	0.0049			-0.0035	0.0035			0.0050	0.0035
Unemrate			0.0157	0.0421			***0.0741***	0.0383			-0.0076	0.0345
Constant	-0.3758	0.3725	-0.2287	1.1636	**-1.4983**	0.1617	**-2.1942**	0.8608	**-1.1530**	0.1790	**-2.5772**	0.8202
Number of obs	1008		844		1825		1377		1829		1492	
LR chi2(16)	206.76		178.86		477.74		411.74		558.79		453.76	
Prob> chi2	0		0		0		0		0		0	
Log likelihood	-374.143		-326.171		-933.327		-683.196		-792.492		-639.499	
Pseudo R2	0.2165		0.2152		0.2038		0.2316		0.2607		0.2619	

References

Baker, S.G. and Laird N.M. (1988), "Regression analysis for Categorical Variables With Outcome Subject to Nonignorable Nonresponse", Journal of the American Statistical association, 83, 401, 62-69.

Baker, S.G. (1995), "Marginal Regression for Repeated Binary Data with Outcome Subject to Non- ignorable Non-response", Biometrics, 51, 1042-1052.

Bring, J. and Carling, K. (2000), "Attrition and Misclassification of Drop-outs in the Analysis of Unemployment Duration", Journal of Official Statistics, 16, 4, 321-330.

Chambers, R.L. and Welsh, A.H. (1993), "Log-linear Models for Survey Data with Non-ignorable Non-response", 55, 1, 157-170.

Cochran, W. and Rubin, D.B. (1973), "Controlling Bias in Observational Studies", Sankyha, 35, 417-446.

Conaway, M.R. (1993), "Non- ignorable Non-response Models for Time-ordered Categorical Variables", 42, 1, 105-115.

Dehejia, R.H and Wahba, S. (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programmes", Journal of the American Statistical Association, 94, 1053-1062.

Gelman. A and Rubin, D. (1992) 'Inference from Iterative Simulation Using Multiple Sequences', Statistical Science, vol.7, pp.457-511.

Hausman and Wise (1979) 'Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment', Econometrica, 47, pp.97-109.

Heckman, J.J., Ichimura, H. and Todd, P.E. (1997), "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", Review of Economic Studies, 64, 605-654.

Heckman, J.J., Ichimura, H. and Todd, P.E. (1998), "Matching as an Econometric Evaluation Estimator", Review of Economic Studies, 65, 261-294.

Heckman, J.J., Ichimura, H., Smith, J.A. and Todd, P. (1998), "Characterising Selection Bias Using Experimental Data", Econometrica, 66, 5

Heckman, J.J., LaLonde, R.J., Smith, J.A. (1998), "The Economics and Econometrics of Active Labour Market Programmes", in Ashenfelter, O. and Card, D. (eds.), The Handbook of Labour Economics, Volume III.

Hellerstein, J.K. and Imbens, G.W. (1999), "Imposing Moment Restrictions from Auxiliary Data by Weighting", The Review of Economics and Statistics, 81(1), 1-14.

Hirano, K, Imbens, G, Ridder, G, and Rubin, D. (1998) 'Combining Panel Data Sets with Attrition and Refreshment Samples', NBER working paper.

Hirano, K, Imbens, G, Ridder, G, and Rubin, D. (2001) 'Combining Panel Data Sets with Attrition and Refreshment Samples', Forthcoming Econometrica.

Horowitz J.L. and Manski, C.F. (1998), "Censoring of outcomes and regressors due to survey nonresponse. Identification and estimation using weights and imputations", Journal of Econometrics, 84, 37-58.

Imbens, G. (2000), "The Role of Propensity Score in Estimating Dose-Response Functions", Biometrika, 87, 3, 706-710.

Imbens, G.W. and Ridder, G. (2000), “Semiparametric Estimation of Panel Data Models with Selective Attrition and Refreshment Samples”, mimeo.
Kish, L. and Hess, I. (1959), “A ‘Replacement’ Procedure for Reducing the Bias of Nonresponse”, American Statistician, 13, 17-19.
Lechner, M. (2001), Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption, in: Lechner, M., Pfeiffer, F. (eds), Econometric Evaluation of Labour Market Policies, Heidelberg: Physica/Springer, p. 43-58.
Little, R. and Rubin, D. (1987) ‘Statistical Analysis with Missing Data’, New York, Wiley.
Ridder, R. (1992), “An Empirical Evaluation of Some Models for Non-Random Attrition in Panel Data”, Structural Change and Economic Dynamics, 3, 2, 337-355.
Robins, J.M., Rotnitzky, A. and Zhao L.P. (1995), “Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data”, 90, 429, 106-121.
Rosenbaum, P.R. and Rubin, D.B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika, 70, 1, 41-55.
Rosenbaum, P.R. and Rubin, D.B. (1983), “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome”, Journal of the Royal Statistical Society, 45, 2, 212-218.
Rosenbaum, P.R. and Rubin, D.B. (1984), “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score”, Journal of the American Statistical Association, 79, 387, 516-524.
Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score", The American Statistician, 39, 1, 33-38.
Rotnitzky, A, Robins, J.M. and Scharfstein, D.O. (1998), “Semiparametric Regression for Repeated Outcomes With Nonignorable Nonresponse”, 93, 444, 1321-1339.
Rubin, D.B. (1974), "Estimating Causal Effects of Treatments in Randomised and Non-Randomised Studies", Journal of Educational Psychology, 66, 688-701.
Rubin, D.B. (1997), “Assignment to Treatment Group on the Basis of a Covariate”, Journal of Educational Statistics, 2, 1, 1-26.
Rubin, D.B. (1979), “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies”, Journal of the American Statistical Association, 74, 366, 318-328.
Rubin, D.B. (1980), "Bias Reduction Using Mahalanobis-Metric Matching", Biometrics, 36, 293-298.
Rubin, D.B. and Zanutto, E. (2001), Using Matched Substitutes to Adjust for Nonignorable Nonresponse through Multiple Imputations”, Survey Nonresponse (Ed), Ch.26.
Tanner and Wong (1990) ‘The Calculation of Posterior Distributions by Data Augmentation’, Journal of the American Statistical Association, vol.82, no. 398, pp.528-550.
Todd, P. (1999), “A Practical Guide to Implementing Matching Estimators”, mimeo.

Table A1: Variable Definitions

Variable	Definition	Values
Labour Market State		
Y0	Good labour market state at period 0, on leaving school	0,1
Y1	Good labour market state at period 1, one year after leaving school	0,1
Y1*	Good labour market state at period 1, one year after leaving school as reported at end of period 1 at second sweep.	0,1
Y2	Good labour market state at period 2, 2 years after leaving school	0,1
	Where:	
	0- Unemployed, Govt Training Scheme, 1- Work FT, Education	
Basic Personal Characteristics		
Gender	1- Male, 0- Female	0,1
Ethnic	0- White, 1 - Non-white	0,1
Home/Family Characteristics		
Fathqual	1- If Father has A-level or higher education, 0 -Otherwise	0,1
Mothqual	1- If Mother has A-level or higher education, 0 -Otherwise	0,1
Pared	1 - If either parent has A-level or higher	0,1
Mothjob	1 - If Mother is in full time work	0,1
Fathjob	1 - If Father is in full time work	0,1
Siblings	Number of Brothers and Sisters	Integer 0-8
Home	1 - If person lives in owner occupied home, 0 - Otherwise	0,1
Livepare	1 - I person lives with one or both parents, 0 - Otherwise.	0,1
Individual's School Performance		
English	Obtained grade A-C GCSE English	0,1
Maths	Obtained grade A-C GCSE Maths	0,1
Gcse	GCSE Score in points	0- 80
Gcsediff	Difference of person score from average school GCSE score	-39.8-56.8
Truant	1 - If person regularly truanted from school, 0 -otherwise	0,1
Individual's Labour Market Anticipation		
Careers	1 - If person had a Careers interview at school, 0 -otherwise	0,1
Workexp	1 - If person had work experience at school, 0 -Otherwise	0,1
School Characteristics		
Avcl1415	Average class size for age 14/15	14.6-29
Specwpct	% of Children with Special Educational Needs	0-100
Permexcl	% of Children Permanently Excluded from the School.	0-0.028
Average_	Average GCSE points score in the school	0.5-68.7
LEA Exam Performance		
Leaavpts	Average GCSE points score in the LEA	31.5-39.3
Locality Characteristics		
Depind	Local Deprivation Index of ward of school	5.0-71.98
Unemrate	Unemployment rate in the ward	3.0-19.5
Predicted Attrition Probability		
Score	Propensity Score	0-1

Table A2: Summary Statistics by Sample

Variable	Balanced Panel and Incomplete Panel BP&IP			Balanced Panel BP			Incomplete Panel IP			Refreshment Sample			BP+R Matched			BP+R/I Matched			R/I Matched			R Matched		
	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.
Labour Market State																								
Y0	1829	0.743	0.437	1008	0.795	0.404	821	0.680	0.467	880	0.342	0.475	1825	0.6729	0.4693	1829	0.7086	0.4545	821	0.6029	0.4896	817	0.5226	0.4998
Y1	1829	0.737	0.440	1008	0.793	0.406	821	0.669	0.471	880	--------	--------	1825	0.4378	0.4963	1829	0.6501	0.4771	821	0.475	0.4997	817		
Y1*	1829	--------	--------	1008	0.813	0.389	821	--------	--------	880	0.299	0.458	1825	0.6707	0.4701	1829	0.6703	0.4702	821	0.553	0.4975	817	0.4945	0.5003
Y2	1829	--------	--------	1008	0.818	0.386	821	--------	--------	880	0.274	0.446	1825	0.6581	0.4745	1829	0.7272	0.4455	821	0.6151	0.4869	817	0.4602	0.4987
Basic Personal Characteristics																								
Gender	1829	0.535	0.499	1008	0.572	0.495	821	0.490	0.500	880	0.343	0.475	1825	0.5775	0.4941	1829	0.5517	0.4975	821	0.5262	0.4996	817	0.5838	0.4932
Ethnic	1829	0.024	0.152	1008	0.031	0.173	821	0.015	0.120	880	0.040	0.196	1825	0.1025	0.3033	1829	0.0284	0.1662	821	0.0256	0.158	817	0.1909	0.3933
Home/Family Characteristics																								
Fathqual	1829	0.214	0.410	1008	0.237	0.426	821	0.185	0.389	880	0.143	0.350	1825	0.2268	0.4189	1829	0.2083	0.4062	821	0.173	0.3784	817	0.2142	0.4105
Mothqual	1829	0.192	0.394	1008	0.215	0.411	821	0.164	0.371	880	0.132	0.338	1825	0.2088	0.4065	1829	0.1974	0.3981	821	0.1754	0.3805	817	0.2007	0.4008
Pared	1829	0.306	0.461	1008	0.331	0.471	821	0.275	0.447	880	0.217	0.412	1825	0.3238	0.4681	1829	0.3013	0.4589	821	0.2643	0.4412	817	0.3146	0.4646
Mothjob	1829	0.574	0.495	1008	0.595	0.491	821	0.547	0.498	880	0.348	0.477	1825	0.5151	0.4999	1829	0.5566	0.4969	821	0.5091	0.5002	817	0.4162	0.4932
Fathjob	1829	0.698	0.459	1008	0.722	0.448	821	0.669	0.471	880	0.431	0.495	1825	0.5923	0.4915	1829	0.6758	0.4682	821	0.6188	0.486	817	0.4321	0.4957
Siblings	1829	1.169	1.080	1008	1.144	1.041	821	1.200	1.126	880	2.280	1.754	1825	1.5085	1.3845	1829	1.2767	1.2158	821	1.4397	1.3838	817	1.9584	1.606
Home	1829	0.722	0.448	1008	0.775	0.418	821	0.658	0.475	880	0.244	0.430	1825	0.657	0.4748	1829	0.6976	0.4594	821	0.6029	0.4896	817	0.5116	0.5002
Livepare	1829	0.967	0.180	1008	0.979	0.143	821	0.951	0.215	880	0.619	0.486	1825	0.9101	0.2861	1829	0.9377	0.2418	821	0.8867	0.3171	817	0.825	0.3802
Individual's School Performance																								
English	1829	0.617	0.486	1008	0.696	0.460	821	0.519	0.500	880	0.213	0.409	1825	0.6258	0.4841	1829	0.626	0.484	821	0.5396	0.4987	817	0.5386	0.4988
Maths	1829	0.470	0.499	1008	0.548	0.498	821	0.375	0.484	880	0.117	0.322	1825	0.4433	0.4969	1829	0.4817	0.4998	821	0.4007	0.4903	817	0.3146	0.4646
Gcse	1829	34.217	17.881	1008	37.519	17.143	821	30.162	17.944	880	13.581	16.022	1825	32.38	18.272	1829	33.88	18.43	821	29.413	18.974	817	26.04	17.627
Gcsediff	1702	4.192	17.174	927	1.114	16.564	775	7.873	17.175	617	17.681	16.684	1524	4.0198	17.049	1646	3.1417	16.921	719	5.7561	17.029	597	8.5317	16.821
Truant	1829	0.382	0.486	1008	0.323	0.468	821	0.454	0.498	880	0.320	0.467	1825	0.3189	0.4662	1829	0.3253	0.4686	821	0.3276	0.4696	817	0.3133	0.4641
Individual's Labour Market Anticipation																								
Careers	1829	0.878	0.327	1008	0.900	0.300	821	0.851	0.356	880	0.538	0.499	1825	0.8153	0.3881	1829	0.865	0.3419	821	0.8222	0.3826	817	0.7111	0.4535
Workexp	1829	0.903	0.296	1008	0.903	0.296	821	0.903	0.297	880	0.689	0.463	1825	0.869	0.3374	1829	0.8814	0.3235	821	0.8551	0.3523	817	0.8274	0.3781
School Characteristics																								
Avcl1415	1602	21.801	2.314	868	21.656	2.375	734	21.972	2.229	565	21.457	2.027	1408	21.535	2.192	1530	21.643	2.3209	662	21.625	2.2491	540	21.339	1.845
Specwpct	1761	1.486	2.717	960	1.316	1.168	801	1.691	3.812	636	5.703	19.718	1564	3.1723	12.868	1695	2.1408	8.6366	735	3.2184	12.973	604	6.1228	20.318
Permexcl	1731	0.003	0.003	948	0.003	0.003	783	0.003	0.003	628	0.005	0.004	1544	0.0033	0.0035	1676	0.0029	0.0032	728	0.0032	0.0035	596	0.0044	0.004
Average_	1702	38.329	8.444	927	38.509	9.069	775	38.114	7.629	617	32.695	9.518	1524	36.796	9.8389	1646	37.923	9.3869	719	37.168	9.7357	597	34.136	10.389
LEA Exam Performance																								
Leaavpts	1702	36.621	2.681	927	36.095	2.710	775	37.251	2.507	605	34.950	3.073	1507	35.715	2.8332	1639	35.925	2.8012	712	35.704	2.9031	580	35.107	2.9212
Locality Characteristics																								
Depind	1791	31.397	14.355	981	31.427	14.281	810	31.360	14.451	659	36.175	14.812	1615	32.281	14.096	1726	31.976	14.384	745	32.698	14.495	634	33.601	13.711
Unemrate	1803	8.691	4.842	991	8.390	4.736	812	9.059	4.946	657	9.699	4.412	1610	8.1249	4.6338	1739	8.4276	4.6783	748	8.478	4.6036	619	7.7012	4.4366
Predicted Attrition Probability																								
Score	1829	0.448	0.184	1008	0.387	0.164	812	0.524	0.179				1825	0.4498	0.1308	1829	0.4491	0.1315	821	0.4877	0.135	817	0.4894	0.1332

Table A3: Summary Statistics for YCS and Matched Refreshment Sample.

Sample	Y0	Y2	W	No of obs
Balanced Panel	0	0	1	108
N_{BP} =1008	0	1	1	99
	1	0	1	75
	1	1	1	726
Incomplete	0	-	0	263
Panel N_{IP} =821	1	-	0	558
Matched	0	0	-	327
Refreshment	0	1	-	63
Sample	1	0	-	114
N_R =817	1	1	-	313

Appendix B: Obtaining the ROUTES 502, Targeted Refreshment Sample in the North East of England

The collection of data was mainly carried out by researchers in the Education Department at Newcastle University. This research team has had extensive experience through a series of recent research projects in the North-East in identifying and contacting young people characterised as disaffected and disadvantaged. That experience was invaluable in locating, '… low achievers and potential drop outs' from national Youth Cohort Surveys'. Out strategy was to contact the voluntary organisations, housing projects, sheltered accommodation and training providers with whom we had worked previously, then, acting on advice from them, to contact mainstream statutory and voluntary organisations as well as smaller and less well known less well known charities, and local groups. It is only by using these 'gatekeepers' could we get access to all the young people who had problems associated with truancy, housing, drugs, crime, families or simply the school to work or child to adult transition process with a lack of family and school support. A list of the 46 agencies who we worked with is attached below.

Contact was usually made with the gatekeeper organisation firstly by letter, followed by a 'phone call, generally followed by a personal visit to talk about the ROUTES project in some detail. Agency managers and workers would then discuss our requirements and suggest names of some of their clients who might be appropriate. When these were approved the agencies would then arrange interview times for us. In some centres members of the ROUTES team also made presentations to groups of potential interviewees to seek their co-operation.

List of Organisations which helped us to locate and interview ROUTES 502 sample:

1	Aquilla Housing Association
2	Barnados Training
3	Benwell Action Research Project
4	Big Lamp Youth Project
5	Blakelaw Integrated Youth Project
6	Byker YMCA Detached Youth Project
7	Cruddas Park Youth Challenge
8	Cruddas Youth Project
9	Daisy Hill Youth Centre
10	Elswick Girls Project
11	Gaskell Avenue Detached Youth Project
12	Gateshead Careers Club
13	Hebburn Detached Youth Project
14	Kenton Young Women's Group
15	Monkchester Young Peoples Project
16	Murray House Initiative
17	National Association for the Care and Settlement of Offenders NACRO Wallsend
18	NEETA Training (Gateshead)
19	Newbiggin Hall Detached Youth Work Project
20	Newcastle Careers Club
21	Newcastle Independence Network
22	Newcastle Social Services
23	North Tyneside MBC
24	North Tyneside Motor Project
25	North Tyneside Training Services
26	Park Avenue Detached Youth Project
27	Pathways to work
28	Rathbone Community Industry, Wallsend
29	Scotswood Youth Strategy
30	South Shields Detached Youth Project
31	South Tyneside Careers Club & Bridge Programme
32	South Tyneside MBC
33	Stepping Stones
34	Streets Ahead Detached Youth Project
35	Streetwise
36	Team Valley Skills
37	The Base
38	The DePaul Trust
39	Tyneside Foyer
40	Walker YMCA
41	West End Youth Enquiry Service
42	YMCA Detached Youth Project
43	Young People's Services
44	Youth Enquiry Service
45	Youth Information Shop
46	Zodiac Training