Examining Ways to Handle Non-Random Missingness in CEA through Econometrics and Statistics Lenses

Jackie Yenerall, Virginia Tech, jyen15@vt.edu
Wen You, PhD, Virginia Tech, wenyou@vt.edu
George Davis, PhD, Virginia Tech, georgedavis@vt.edu
Paul Estabrooks, PhD, Virginia Tech, estabrkp@vt.edu

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.

1 Introduction

Missing values which result from attrition are common in randomized trials. Failure to appropriately address the missingness could result in biased estimates and misleading conclusions. This is of particularly concern in cost-effectiveness analysis (CEA) as any bias in either the point estimates for cost or effects will then bias the cost-effectiveness estimate. Since these cost-effectiveness estimates are used to inform health policy, it is necessary to be especially careful regarding how missing data is handled (Faria et al, 2014). Complicated experimental designs, such as cluster randomized trials (CRT) and trials which follow individuals over time (longitudinal or panel data), also call for greater care when addressing missingness.

Of greatest concern is when observations are systematically missing, i.e. missing not at random. Sample selection models are one method used to address potential bias resulting from this mechanism of missingness. The purpose of this paper is to compare two sample selection models used to address non random missing values and determine the relative advantages of either when applied to a longitudinal cluster randomized trial. From the economics literature we will utilize the Heckman (1976) sample selection model and the Diggle and Kenward (1994) model from the statistics literature. The analysis will utilize both data collected as part of a cluster randomized trial for a worksite weight loss program, and a simulation study. This paper presents preliminary results which analyzed the data from the CRT trial, and will propose the design for the simulation study.

2 Current Practice

Analyzing data from a cluster randomized trial (CRT) in a cost effectiveness analysis (CEA) presents many challenges. Data from a CRT is complicated by the presence of grouping of individual observations in levels such as hospitals, worksites, classrooms, etc. Such grouping, or clustering as it is also known, is problematic because it can induce correlation in the error term, and result in biased estimation (Goldstein, 1995). Further complications are introduced when trying to conduct a CEA given the bivariate nature of the outcomes (cost and effects) which may be correlated, and for which cost may have a skewed distribution (Gomes et al, 2011; Flynn and Peters, 2005). There has been substantial progress on identifying the best practices for handling these challenges as they specifically relate to the CEA of data from a CRT, however, what receives less attention is addressing missing data (Gomes et al, 2011; Gomes et al, 2013).

Missing data is a routine occurrence in experiment data, and if not addressed appropriately it can bias estimations (Nobel et al, 2012; Faria et al., 2014; Rubin and Little, 2002; Diaz-Ordaz et al., 2014). Often the missingness is a result of attrition, which occurs when experiments follow individuals over time and participants are either lost to follow-up or drop out of a study. Previous literature has consistently found that when data is missing either in CEA, CRT, or CEA of CRT data it is most often ignored, or simplistic methods such as complete case analysis or last observation carried forward are utilized (Dias-Ordaz et al, 2014; Dias-Ordaz et al, 2014a; Dias-Ordaz et al, 2014b; Nobel et al, 2012). These methods are valid only when observations are missing completely at random, which is very difficult to justify in a clinical trial (Rubin and Little, 2002; Dias-Ordaz et al, 2014b).

Within the literature on missing data in CEA and CRT the focus has been on adapting multiple imputation (MI) to address the complex nature of the data. However, this method is most often used to address cases when it is assumed the data is missing at random (MAR), although it can be adapted for use in the missing not at random. case (Ruin, 1996; Gomes et al, 2013; Diaz-Ordaz et al, 2014a). Previous research on the application of MI to CEA of CRT and has shown that it is important to incorporate the hierarchical structure of the dataset into the method used to address missingness to avoid bias estimation (Gomes et al, 2013; Diaz-Ordaz et al, 2014a). To do this, researches have used multilevel models in both the imputation and analysis stages of MI (Gomes et al, 2013; Diaz-Ordaz et al, 2014a). However, these models were only designed to analyze missing data that resulted from MAR and from a two time period study. In one application, it was shown that when the true missingness mechanism was MNAR the MI model adjusted for the hierarchical nature of the data was still biased (Gomes et al, 2013). This clearly demonstrates the importance of identifying models that will yield valid inference under the MNAR mechanism and are suited to handle the complex nature of CEA and CRT data.

Thus it is important to assess the ability of various methods to appropriately handle not only the complicated nature of CRT data, but also the mechanism of missingness in the data. Sample selection models are one option when trying to address potential bias from attrition. Two popular forms of this model are the Diggle and Kenward model from the statistics literature and the Heckman model from econometrics. The purpose of this paper will be to investigate their appropriateness when utilized in the analysis of cost effectiveness data from a cluster randomized trial.

3 Comparison of Sample Selection Models

Both the Heckman and Diggle Kenward (DK) sample selection models contain two parts: an outcome mechanism and a selection mechanism (Diggle and Kenward, 1994; Heckman, 1979) . The outcome mechanism describes the process generating the outcome of interest, as if it were fully observed. However, with attrition estimating this process alone could result in bias. To address this bias, both sample selection models incorporate a sample selection mechanism, which describes the process which results in attrition. By estimating both the outcome and selection mechanisms, the models hope to correct for the bias in the outcome mechanism.

A major difference between the Heckman and DK model comes in the specification of the sample selection equation. While the DK model relies upon trends in the outcome to identify the selection process, the Heckman model relies on a latent process, which is correlated with the outcome of interest. To further illustrate the differences we will now review each model in more depth.

3.1 Diggle Kenward Sample Selection Model

An important component of models for missingness in the statistics literature is the notion of mechanism of missingness which were first developed by Rubin (1976), in a paper in which he sought to determine the conditions under which missingness could be ignored without biasing the analysis of the outcome of interest.

To introduce these mechanisms, we first need to establish some notation. Since we are working with longitudinal data, the index $i = 1, ...N$ refers to the number of individuals, and the index $j = 1, ...T$ refers to the number of time periods under observation. Let $Y^* = (y^*_{ij})$ denote the complete sample that would have been obtained if there was no missing data. Let the missing components of Y^* be noted as missing $Y^m = (y^m_{ij})$ and the observed $Y^o = (y^o_{ij})$ (Diggle et al., 2002). The random variable indicator for drop out, D_i, used to indicate this partition in the data set. Thus, in the case that drop out occurs at $D_i = d$ then $y_{i1}, ...y_{i,d-1}$ are observed and $y_{id}, ...y_{iT}$ are missing.

Combining the partition of the outcome and the random variable for missingness, results in the joint distribution of the sample: $f(Y^o, Y^m, D|\theta, \beta)$ where θ and β are unknown parameters for the distribution of the outcome, Y, and missingness, D, respectively. Using standard factorization rules we can rewrite it as:

$$f(Y^o, Y^m, D|\theta, \beta) = f(Y^o, Y^m|\beta) f(D|Y^o, Y^m, \theta) \tag{1}$$

To make this distribution compatible with likelihood methods, it is then necessary to integrate out the missing values so that the joint distribution is in terms of the observed random variables only (Diggle et al., 2002).:

$$f(Y^o, D|\theta, \beta) = \int f(Y^o, Y^m|\beta) f(D|Y^o, Y^m, \theta) dY^m \tag{2}$$

This particular form of factorization results in the sample selection model, where $f(Y^o, Y^m|\beta)$ describes the outcome process, and $f(D|Y^o, Y^m, \theta)$ describes the selection or attrition or missingness process. In the case that these processes are distinct (i.e. the parameter space of (θ, β) is the product of the parameter space for θ and β) then the missing-data mechanisms is ignorable, which implies that the outcome process can be estimated independently of the missingness process without bias (Rubin and Little, 2002). Determining distinctness relies upon identifying the missingness mechanism in the data. The classifications of missingness mechanisms were first introduced by Rubin and Little (1987) and focuses on further defining the conditional distribution of missingness, based on the relationship between the random variable for missingness and the outcome (Rubin and Little, 2002). For panel datasets, this requires defining the relationship between past and current observations.

The most restrictive case is called missing completely at random (MCAR), this assumes that the missingness random variable is independent of both the observed and missing outcome thus Rubin and Little, 2002):

$$Pr(D_i = d|y_{i1}, ..., y_{iT}) = \theta \text{for all} y_{i1}, ..., y_{iT} \tag{3}$$

In the case that the missingness variable depends on the observed outcome, but not the missing outcome it's known as missing at random (MAR). In the case of panel data, this implies that missingness can depend on values observed prior to drop out, but not on values at the time of drop out (Rubin and Little, 2002).

$$Pr(D_i = d|y_{i1}, .., y_{iT}; \theta) = Pr(R_t = d|y_{i1}, .., y_{i,d-1}, \theta)\text{for all} y_{i1}, ..., y_{iT} \tag{4}$$

Finally, in the case that missingness depends on both the observed and missing observation, the conditional distribution cannot be further simplified and results in the missing not at random (MNAR) mechanism. From the above discussion, it can be seen that the missingness mechanism is only ignorable in the MCAR and MAR case. In the case we have reason to believe that the data follows a MNAR mechanism, it will be necessary to estimate the full sample selection model.

The Diggle and Kenward model is one of the most widely cited examples of how to operationalize the sample selection model for the case of attrition in the panel data setting (Diggle and Kenward, 1994; Rubin and Little, 2002; Molenberghs and Kenward, 2007). The purpose of their original paper was to develop a model that could accommodate MCAR and MAR as special cases in an MNAR model (Diggle and Kenward, 1994). Diggle and Kenward noted in their paper that they believe developing a model that could distinguishing and test for differences between MNAR and MAR was a more important than distinguishing between MCAR and MAR (which had been the focus of previous papers) because failure to distinguish between the former would result in bias parameter estimates of the outcome equation (Diggle and Kenward, 1994). In the case that the data are MNAR, individuals who are observed differ systematically from those who are missing, this implies that the conditional expectation of Y^* is different from Y^o (Diggle and Kenward, 1994). More importantly, they believe that different trends amongst the observed and unobserved individuals drives this bias. Thus, by incorporating this theory into their selection process, and jointly estimating the selection and outcome process they can avoid biased estimates.

To estimation this selection model, we need to make assumptions about the selection process ($f(R|Y^o, Y^m, \theta)$ in our previous notation), and the marginal distribution of the outcome ($f(Y^o, Y^m|\beta)$ in our previous notation). One of the most important assumptions in their model, is that if an individual is still observed at time $j = k$, the sequence of measurements ($Y = y_{i1}, ..., y_{ik}$) follows the same as that for the complete outcome ($Y^* = y^*_{i1}, ..., y^*_{ik}$), (Diggle and Kenward, 1994). In their model, the sample selection process is a function of the history of observed observations up to the time of drop out at $j = d$, $H_d = (y_{i1}, ..., y_{i,d-1}$, and y^*_{id} the value of the outcome that would have been observed had the individual not dropped out. The probability of drop out is given by:

$$P(R = d|H_d) = p_d(H_d, y^*_{id}; \theta) \tag{5}$$

Given that this model was only for the attrition pattern, it assumes there is no missingness in the first time period and that:

$$P(Y_k = 0|H_k, Y_{k-1} = 0) = 1 \tag{6}$$

This formulation of the probability of missingness (i.e. the sample selection process) includes the ability to distinguish between the missingness mechanism discussed earlier, where MNAR and MAR result from different assumptions about the relationship between missingness and the unobserved observations. In the case of MCAR, $p_d()$ depends on neither H_d nor y^*_{id}; in the car of MAR, $p_d()$ depends on H_d but not y^*_{id}; and finally in MNAR, $p_d()$ depends, most importantly, on y^*_{id} and possibly also on H_d.

The distribution of the outcome, y^*_{ij}, is represented by $f^*_{ij}(y_{ij}|H^*_j;\beta)$ where $H^*_j = i^*_{i1},...,y^*_{i,j-1}$ and is normally distributed, and , while the conditional distribution of y_{ij} is represented by $f_{ij}(y|H_j;\beta)$, where $H_j = y_{i1},...,y_{i,j-1}$. Combining all of this, we can then write the distribution for completers (i.e. those who do not drop out), suppressing the dependence on the parameters θ, β :

$$\begin{aligned} f(y) &= f^*_1(y_1)\prod_{k=2}^{T} f_k(y_k|H_k) \\ &= f^*(y)\left(\prod_{k=2}^{t}(1-p_k(H_k,y_k))\right) \end{aligned} \tag{7}$$

And for those who experience drop out at time $j = d$:

$$\begin{aligned} f(y) &= f^*_1(y_1)\left(\prod_{k=2}^{d-1} f_k(y_k|H_k)\right)P(Y_d=0|H_d) \\ &= f^*_{d-1}(y^{d-1})\left(\prod_{k=2}^{d-1}(1-p_k(H_k,y_k))\right)P(Y_d=0|H_d) \end{aligned} \tag{8}$$

Combining across the full population (ie. completers and drop outs) yields a log-likelihood function which can be written in the following partitioned form:

$$
\begin{aligned}
&L(\theta,\beta) = L_1(\beta) + L_2(\theta) + L_3(\theta,\beta) \\
&\text{Where:} \\
&L_1(\beta) = \sum_{i=1}^{N} log(f^*_{d-1}(y_i) \\
&L_2(\theta) = \sum_{i=1}^{N}\sum_{k=1}^{d_1-1} log(1 - p_k(H_k, y_{ik}) \\
&L_3(\theta,\beta) = \sum_{i:d_i\leq n} log(Pr(R = d_i|y_I))
\end{aligned} \tag{9}
$$

To estimate this model, they assume that missingness follows a logistic distribution: $logit[p_k(H_k, y; \theta)] = \theta_0 + \theta_1 y + \sum_{j=2}^{k} \theta_j y_{k+1-j}$. However, in most applications the mechanism is restricted to only consider the current time period in which an individual drops out, and the previous period result in the following specification: $logit[p_k(H_k, y; \theta)] = \theta_0 + \theta_1 y_{i,j-1} + \theta_2 y_j$. This formulation of the missingness mechanism also provides the basis for their test of MAR vs MNAR, which utilizes the earlier discussion of the application of the mechanisms to panel data. In panel, the missingness mechanism is MAR if it depends on all past observations, but not the current missing observation, while MNAR depends on the current observation and also the past. By making their missingness mechanism a function of the history of the outcome they were able to incorporate both mechanism in such a way that allowed for testing between the two. To test for an MNAR mechanism consider the coefficient θ_2 which is associated with y_{ij}, the observation for the current time period, which may be missing. In the case that θ_2 is significantly different from zero, it would imply the presence of an MNAR mechanism.

The outcome model follows a multivariate normal distribution, with a variance covariance structure that accounted for serial correlation. In the terminology of econometrics, their outcome model follows a random effects model (estimated using maximum likelihood) with adjustment for serial correlation. Given the presence of unobserved random effects, and missing data the log likelihood is highly complex, with multiple high dimensional integrations (for both missing data and the random effects), and can be difficult to estimate in the presence of high rates of attrition.

One criticism of this model, and sample selection models in general, is that they are sensitive to model specification, especially in the sample selection process (Little and Rubin, 2002; Verbeke et al, 2001; Molenberghs and Kenward, 2007). In the words of Diggle and Kenward (2002) the sample selection process "conveys the notion that dropouts are selected according to their measurement history," while they attribute this interpretation of the sample selection process to Heckman (1976) we will see next that this interpretation misses a key assumption about the sample selection model in econometrics, which is that the sample selection process is driven by a separate mechanism (Heckman, 1976).

3.2 Heckman Sample Selection Model

Sample selection models were originally developed in the cross sectional setting for cases in which a large point mass was observed over zero, and it was believed that using just the positive values would bias the relationship of interest (Amemiya, 1985). In Heckman's (1976) seminal paper on the topic, the interest was in estimating the wage function for women, however, a large number of observations have a recorded wage of zero in the survey. Their observations are not missing, rather the zero values reflect their choice not to work. Thus, to determine the wage equation it must be jointly considered with the individuals choice to work. In the case that there are unobserved factors which influence both the decision to work and wages, then estimation using ordinary least squares will be biased. Here, unlike in the Diggle and Kenward model, the bias is driven by omitted variables that are present in both the sample selection process and the outcome. Given the theorized presence of the same omitted variables in both equations, the two equations will be correlated requiring their joint estimation.

Heckman (1976) noted that this model can be applied to missing data which results in censoring and can also be applied to missing data (Heckman, 1976; Amemiya, 1985). Instead of choosing to work, individuals now chose to be part of the sample. The model is represented by the following set of equations:

$$\begin{aligned} R &= 1 \text{if} R^* > 0 \\ &= 0 \text{if} R^* \leq 0 \\ Y^* &= Y^o \text{if} R = 1 \\ &= Y^m \text{if} R = 0 \end{aligned} \tag{10}$$

Where:

$$
\begin{aligned}
R^* &= W\theta + V \\
Y &= X\beta + U
\end{aligned}
\tag{11}
$$

In this formulation of the sample selection model, both the outcome and sample selection process follows a latent variable process (where R^*, Y^* represents the latent variables). This implies that while we only observe the binary outcome R, the process of selecting into the same actually follows a continuous process. This latent variable process represents the individuals optimization problem, in the case that they determine it is optimal to be observed in the sample $R^* > 0$, $R = 1$ and Y is observed.

The goal of the model is to estimate the outcome equation, Y^*, however, as was discussed previously Y^* is only observed in the case that that $R = 1$. The question then is, can we consistently estimate Y^* given the sample selection process?

$$
E(Y|X, R \geq 0) = X\beta + E(U|V \geq -W\theta) \tag{12}
$$

Clearly, unless U is independent of V, the conditional mean of W will be non zero, and a sample selection bias in the parameters of Y will result. Given the linear form of both the selection process and the outcome, assuming a joint distribution (here bivariate normal) for the error terms allowed Heckman to derive the exact form of the sample selection bias

(Heckman, 1976).

$$\begin{pmatrix} V \\ U \end{pmatrix} \sim N\{\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & \sigma_{VU} \\ \sigma_{VU} & \sigma_U^2 \end{pmatrix}\} \tag{13}$$

Since the error terms are bivariate normal, we also know that the error terms will display a linear dependence (Vella, 1998).

$$\begin{aligned} &U|V \sim N(\sigma_{VU}, \sigma_U^2 - \sigma_{VU}^2) \\ &U = \sigma_{VU}V + \epsilon \\ &\text{Where:} \\ &\epsilon \sim N(0, \sigma_U^2 - \sigma_{VU}^2) \end{aligned} \tag{14}$$

We can then substitute this definition for U back into equation **??**:

$$\begin{aligned} &E(Y|X, R = 1) = X\beta + E(\sigma_{vu}V + \epsilon|V > -W\theta) \\ &= X\beta + \sigma_{VU}E(V|V > -W\theta) \\ &\text{Where:} E(V|V > -W\theta) \text{describes the mean of a truncated standard normal} \\ &= W\beta + \sigma_{VU}\left(\frac{\phi(W\theta)}{1 - \Phi(-W\theta)}\right) \\ &= W\beta + \sigma_{VU}\left(\frac{\phi(W\theta)}{\Phi(W\theta)}\right) \\ &= X\beta + \sigma_{VU}\lambda(W\theta) \end{aligned} \tag{15}$$

In the last line, $\lambda(W\theta)$ is known as the inverse mills ratio, and it indicates the probability that an individual will be included in the sample. Although this model was not explicitly designed for longitudinal data it can be used in this setting, it would just assume that the selection process is the same in every period.

One advantage of deriving an explicit form for the bias and utilizing two separate equations for the selection and outcome process is that it increases the options for estimation. Either a two step approach which estimates the sample selection bias, or a maximum likelihood approach which parametrizes the sample selection bias can be used. In either the two step, or MLE the full sample is used to estimate the sample selection equation, and only those who are observed are used in the outcome.

Heckman (1976) suggested the two step method, in which the first stage estimated the selection process using a probit model. Then these results were used to calculate the inverse mills ratio, which was then used as a variable in the second stage ordinary least squares model. Alternatively, a maximum likelihood approach can be used, in which the selection equation and the outcome are jointly estimated, and the correlation between the equations is parameterized in the model as ρ. The log likelihood for the MLE for the Heckman model is as follows:

For those who are observed:

$$L_{ij} = \Phi\left(\frac{w_{ij}\theta + (y_{ij} - x_{ij}\beta\rho/\sigma}{(1-\rho^2)^{1/2}}\right) - 1/2\left(\frac{y_{ij} - x_{ij}\beta}{\sigma}\right)^2 - ln((2\pi\sigma)^{1/2}) \tag{16}$$

For those who are not observed

$$L_{ij} = ln\Phi(-w_{ij}\theta)$$

Much like the Diggle and Kenward model, the Heckman model also has a test for MNAR, or the sample selection bias. In the two step model, this is the same as the test for the significance of coefficient on the inverse mills ratio, and in the MLE it its he test for significance of ρ.

4 Data

This paper uses baseline, 6month, and 12month data from a cluster randomized control trial for worksite weight loss to conduct a cost-effectiveness analysis at 12 months. Recruited worksites were randomly assigned to two groups: a less-intensive quarterly newsletter program focused on providing knowledge on healthful eating and physical activity (LMW); or an individually-targeted internet-based intervention with monetary incentives (INCENT). To be eligible worksites had to: provide internet access to their employees; have between 100 and 600 employees; have employees physically located in one site with access to a central location for kiosk weigh-ins; and agree to conduct a brief health survey of the entire employee population. To be eligible participants had to: be adults ($>$ 18 years old); have a BMI $\geq$ 25 kg/m^2; not currently pregnant or pregnant in the last 12 months; not currently participating in a weight loss program (e.g. Weight Watchers); free of serious medical conditions (e.g. terminal cancer, recent heart attack); be employed by one of the participating worksites; and have access to the internet at their work location. Individuals who withdrew from the intervention at 6 or 12 months due to medical illness or change in employment were excluded from the outcome analysis.

After randomization, 14 worksites were included in the INCENT group and 14 in the LMW group. Baseline statistics were calculated at the worksite level to ensure balance was achieved during randomization. Because the panel is unbalanced (i.e. there are different numbers of participating employees in each worksite) both weighted and unweighted analysis was conducted. The weighted method takes into account the different sized clusters when calculating means and t-tests (Hayes and Moulton, 2009). Since the results of both methods were similar only unweighted means are reported. The similarity of the results provide additional evidence of the validity of the randomization process. Table 1 contains the baseline summary statistics. From the table we can see that at baseline the only significant difference between INCENT and LMW at the cluster level is worksite years (i.e. the number of years an individual was employed at their worksite) and percentage of employees who are managers.

At baseline, 1790 individuals enrolled for the study (1001 in INCENT and 789 in LMW). However, one worksite from LMW dropped out at 6 months and thus participating employees (n=54) from this worksite were excluded from the outcome analysis. Individuals who left the intervention due to serious medical condition or job loss(n=62), were also exclude from analysis. Finally, all individuals missing baseline weights (n=48) were excluded. Thus, there were 27 clusters (14 in INCENT and 13 in LMW) and 1626 individuals (932 in INCENT and 694 in LMW) included in the analysis.

For the cost-effectiveness analysis, change in body mass index (BMI) will be used as the effectiveness outcome. Costs were collected retrospectively (i.e. after the trial ended) and include both the program costs, and the absolute cost of absenteeism. The calculation of program costs differed between the two treatment arms. Program fees for the INCENT group followed the formulation used by a company which currently operates a similar program. The program fee included a monthly base program fee which was incurred per eligible employee, which is an important distinction from the number of participating employees, as well as monthly rental fee for the health spot, travel cost for kick off, and the incentive. The incentive structure was based off percentage weight loss, all relative to the baseline, and incentives were earned at 6 months and 12 months. Incentive amounts started at 5% weight loss, and increased by increments of 5: weight loss of 5-9% earned \$5; 10-14% earned \$10; 15-19% earned \$15, and >20% earned \$20. The total program fee was payed off in installments, with 60% paid at baseline, 20% at 6 months, and the remaining 20% is paid at 12 months.

The LMW program fees were micro-costed since a pre-existing program did not exist upon which to base the fee. Since the LMW was an adaptation of the INCENT program, the base program fee was based off the cost to adapt the newsletter and the mini-sessions. It was assumed that the the payment scheduled was the same for the LMW worksites as it was for the INCENT, thus 40% paid at baseline, 20% at 6 months, and the remaining 20% is paid at 12 months.

Absenteeism is assessed the same for both groups, using the World Health Organization (WHO) Health and Work Performance Questionnaire (HPQ) (Kessler et al, 2003;2004). This survey assesses absenteeism by missed hours of work to account for the variation in work requirements in different occupations (i.e. the "typical" 8 hour work day may not be so typical anymore) (Kessler et al, 2003;2004). Workers are asked not only about the number of hours they work, but also about how many hours they are expected to work and the number of additional hours worked to make up for missed time (Kessler et al, 2003;2004). For this study, absolute absenteeism was utilized, which is calculated as the number of expected hours per month, less the number of hours worked in the past month. Absenteeism was valued using the human capital method, which values time away from work at the wage rate (fringe benefits are also commonly included in implementation). However, this study only collected information on household yearly income at baseline and in interval form. Thus we only have information on which interval a household's income falls into and not their unique income level. For each interval, the midpoint was used to indicate monthly income and was used to value monthly absenteeism.

Table 2 includes the summary statistics on the total costs of the program for the LMW and INCENT group. As with the BMI outcome, the observations associated with the worksite from LMW that dropped out at 6 months (n=54) were excluded from the outcome analysis. Individuals who left the intervention due to serious medical condition or job loss(n=62), and were missing their baseline program costs (n=87) were also exclude from analysis. Thus, at baseline there are 1587 observations (914 in INCENT, 673 in LMW). While there is generally a concern with health costs being skewed, this is generally in reference to direct medical costs, which can be very large for the few individuals who are very ill (Briggs and Gray, 1998; Blough et. al, 1999; Briggs et. al, 2005; Gilleski and Mroz, 2004; Manning etl al, 2005; Manning and Mullahy, 2001; Mihaylova et. al, 2011). However since we do not have information on direct medical costs for this study, we are not concerned with a skewed distribution.

By 6 months, 443 individuals were missing outcome values for the effectiveness outcome, BMI (258 in INCENT and 185 in LMW), resulting in 27.24% attrition (27.68% in INCENT and 26.66% in LMW). By 12 months, 672 individuals (400 in INCENT and 272 in LMW) had missing BMI values, resulting in attrition of 41.33% (42.92% in INCENT and 39.19% in LMW). For costs, by 6 months 454 individuals had missing outcome (277 in INCENT, 177 in LMW) and by 12 months 646 individuals had missing outcomes (387 in INCENT, 259 in LMW).

5 Methods

For the case study portion of this paper we will consider the performance of the Heckman and Diggle Kenward sample selection models when applied to the cost and effectiveness outcomes of the 12 month outcomes for the worksite weight loss program. We also include the results from a linear mixed model that ignores missingness as a comparison to detect bias, but includes random effects at the individual level and controls for autocorrelation.

The results from the linear mixed model are used as starting values for the DK sample selection model with the MCAR missingness mechanism. To estimate the DK sample selection model with MNAR missingness mechanism (the mechanism of primary interest) it is necessary to estimate the MCAR, then MAR and finally the MNAR versions of the model. The results from the MCAR model serve as the starting values for the MAR model, and the results from the MAR model are used as starting values for the MNAR model.

To estimate the Heckman sample selection model we need to specify a selection equation, which includes at least one variable that is not included in outcome model. For these preliminary results we utilize gender and age in the selection equation. In both models, for the outcome equation we use a simple program evaluation specification with a dummy variable for the INCENT group, and treatment interaction terms (time X group indicator) for both INCENT and LMW. Additionally, for the cost data, the outcome had to be estimated in the cost per millions.

6 Results

Since the primary purpose of this paper is to investigate the performance of two sample selection, analysis of the preliminary results will focus on comparing the estimates for the two treatment interaction term (INCENT X Time) from the different models. In the case that significant sample selection bias was present in the worksite data we would expect to see large differences between the point estimates in the linear mixed model, which represents the more naive model that ignores missingness, and the sample selection models which should address bias from attrition.

Table 4 contains the result for the cost data, and from comparing the results from the linear mixed model to those from the MNAR DK model and the Heckman model, we can see there isn't much difference, which would suggest, at an initial glance, that there does not appear to be a sample selection bias in the cost data. Additionally, the DK and Heckman models given similar results. Table 3 contains the results for the effectiveness outcome, where we can see there are more substantial differences both between the linear mixed model and the sample selection models, as well as between the two sample selection models. It should also be noted hat the missing values in the variance covariance matrix in the MNAR model are indication that we have not yet achieved convergence in the DK model. The differences between the linear mixed model and the sample selection models may indicate the presence of selection bias. However, given that the Heckman model relies on linear regression in the outcome, which may be biased when utilized with CRT data a simulation is necessary to further investigate the varied performance between the two estimators.

Additionally, when comparing the performance of the two sample selection models it is also important to consider the ease of estimation. For both costs and effectiveness, the DK model required substantial additional time to estimate as models could take upwards of 15 minutes to achieve convergence. Both also required the convergence criteria to be relaxed in order to achieve convergence. For effects, even after substantially relaxing the convergence criteria it was still not possible to achieve convergence in the MNAR model.

Given the differences in the performance of the two models between the cost and effect outcomes, it suggests the need for a simulation to isolate the effects that different parameters in the cluster randomized trial have on the relative performance of the two models. Following the work of previous simulation papers which evaluated the performance of estimators for CRT in the cross sectional setting, we will include the number of clusters, cluster size, intraclass correlation coefficient (ICC) as parameters in our simulation(Gomes et al, 2011). Given the application to the CEA and longitudinal data, we will also explore the effects of the level of correlation between outcomes, as well as the strength of the trend that occurs in individual outcomes over time.

7 Tables

Table 1: Baseline Summary Statistics, Cluster Level

	Total	INCENT	LMW
BMI	33.32	33.24	33.39
	(1.41)	(1.48)	(1.39)
Severity	5.62	5.57	5.67
	(2.46)	(2.87)	(2.09)
Caucasian(%)	77.32	72.66	81.97
	(21.83)	(20.60)	(22.78)
African American(%)	19.14	24.02	14.26
	(21.75)	(20.99)	(22.15)
Other(%)	3.55	3.32	3.79
	(2.49)	(2.23)	(2.78)
Female(%)	73.82	80.21	67.43
	(20.01)	(10.84)	(25.02)
Male(%)	26.18	19.79	32.57
	(20.01)	(10.84)	(25.02)
Spouse(%)	60.71	58.28	63.14
	(7.94)	(7.36)	(8.00)
Worksite Years	9.75	8.12	11.38**
	(3.37)	(3.00)	(2.98)
Heart Disease(%)	4.71	2.71	6.70
	(8.06)	(3.66)	(10.63)
Diabetes(%)	10.37	8.54	12.21
	(8.89)	(5.67)	(11.17)
High Blood Pressure(%)	32.54	29.04	36.03
	(11.41)	(10.85)	(11.25)
Arthritis(%)	18.34	14.94	21.74
	(10.33)	(5.24)	(13.02)
Income<$50,000(%)	37.93	39.63	36.22
	(16.17)	(16.83)	(15.92)
Income $50,000-$99,999(%)	38.10	37.68	38.53
	(7.56)	(5.53)	(9.36)
Income>$100,000(%)	21.51	21.28	21.73
	(13.72)	(13.98)	(13.99)
Manager(%)	14.14	11.28	17.01**
	(6.02)	(4.27)	(6.27)
Sales(%)	7.56	9.97	5.15
	(14.60)	(18.90)	(8.54)
Service(%)	3.48	3.60	3.35
	(4.14)	(3.96)	(4.46)
Office(%)	19.83	21.63	18.03
	(9.57)	(10.63)	(8.38)
Equipment(%)	7.29	5.66	8.91
	(13.76)	(13.74)	(14.11)
Other(%)	15.07	15.15	14.99

Stars refer to a significant difference between the INCENT and LMW group
*= $p<.05$ **= $p<.01$ ***= $p<.001$

Continued on next page

Table 1–continued from previous page

	Total	INCENT	LMW
	(8.54)	(9.88)	(7.35)

Note: Weighted means were also calculated, but because they were very similar to unweighted means the results are not reported

Stars refer to a significant difference between the INCENT and LMW group
*= $p < .05$ **= $p < .01$ ***= $p < .001$

Table 2: Summary of Costs (INCENT vs LMW)

			Cost Baseline		
mean(sd)	Program Cost	Absenteeism Cost	Incentive Cost	Total Cost	Count
INCENT	64.02	19,258.39		19,323.36	911
	(17.27)	(262,697.48)		(262,695.87)	
LMW	25.33	-1,540.91		-1,515.62	670
	(7.17)	(254,510.76)		(254,510.45)	
Total	47.5	10,444.01		10,492.16	1,581.00
	(23.65)	(259,382.24)		(259,381.94)	
			Cost 6 Months		
	Program Cost	Absenteeism Cost	Incentive Cost	Total Cost	Count
INCENT	32.01	23,165.54	1.09	23,198.97	643
	(8.63)	(281,146.74)	(3.16)	(281,146.01)	
LMW	12.66	15,909.09		15,921.77	500
	(3.58)	(315,522.16)		(315,522.02)	
Total	23.75	19,991.24		20,015.59	1,143.00
	(11.82)	(296,562.76)		(296,562.43)	
			Cost 12 Months		
	Program Cost	Absenteeism Cost	Incentive Cost	Total Cost	Count
INCENT	32.01	12,796.26	1	12,830.03	530
	(8.63)	(257,653.97)	(2.69)	(257,653.58)	
LMW	12.66	26,694.88		26,707.71	419
	(3.58)	(262,390.80)		(262,390.94)	
Total	23.75	18,932.74		18,957.27	949
	(11.82)	(259,710.21)		(259,709.77)	

Table 3: Sample Selection Models: Effectiveness

Coef(SE)		Diggle Kenward			
	Linear Mixed Model	MCAR	MAR	MNAR	Heckman
		Outcome: BMI			
INCENT X Time	-0.91981	-0.8914192	-0.8865487	-0.8868919	-1.6363
	(0.1561592)	(0.1585991)	(0.1585149)	(0.1585111)	(0.2008)
LMW X Time	0.31793	0.3149041	0.3020068	0.3024478	0.546
	(0.1785604)	(0.1806888)	(0.1805895)	(0.1805853)	(0.2192)
		Selection			
$Cost_{j-1}$			0.0273798	2.19E-01	
			(0.00657176)	(1.69E-05)	
$Cost_j$				1.591342	
				(NaN)	
Age					2.11E-05
					(1.46E-05)
Female					-1.36E-01
					(3.70E-02)

Table 4: Sample Selection Models: Costs

Coef(SE)		Diggle Kenward			
	Linear Mixed Model	MCAR	MAR	MNAR	Heckman
		Outcome: Cost/10e5			
INCENT X Time	0.00726	0.00712	0.00707	0.00777	0.00706
	(0.00658)	(0.00665)	(0.00665)	(0.00774)	(0.00744)
LMW X Time	-0.00974	-0.00969	-0.00979	-0.01037	-0.00795
	(0.00752)	(0.00760)	(0.00760)	(0.00827)	(0.00854)
		Selection			
$Cost_{j-1}$			0.16097	0.21116	
			(0.17015)	(0.19494)	
$Cost_j$				0.07599	
				(0.35429)	
Age					2.89E-05
					(1.96E-05)
Female					-1.84E-01
					(4.75E-02)

8 References

Amemiya, Takeshi. "Tobit Models." Advanced Econometrics. Cambridge: Harvard University Press, 1985. 360-411.

Benedict, Michael A., and David Arterburn. "Worksite-Based Weight Loss Programs: A Systematic Review of Recent Literature." *American Journal of Health Promotion* 22.6 (2008): 408-16.

Birmingham, Jolene, Andrea Rotnitzky, and Garrett Fitzmaurice. "Mixture and Selection Models for Analysing Longitudinal Data with Monotone Missing Patterns."*Journal of the Royal Statistical Society. Series B* (Statistical Methodology) 65.1 (2003): 275-97.

Blough, David , Carolyn Madden, and Mark Hornbrook. "Modeling risk using generalized linear models." *Journal of Health Economics* 18 (1999): 153-171.

Briggs, Andrew, and Alastair Gray. "The distribution of health care costs and their statistical analysis for economic evaluation ." *Journal of health services research* & *policy* 3.4 (1998): 233-245.

Briggs, Andrew, Richard Nixon, Simon Dixon, and Simon Thompson. "Parametric modelling of cost data: some simulation evidence." *Health Economics* 14 (2005): 421-428.

Diaz-Ordaz, Karla, Michael Kenward, Abie Cohen, Claire Coleman, and Sandra Eldridge. "Are Missing Data Adequately Handled in Cluster Randomized Trials? A Systematic Review and Guidelines." *Clincal Trials* (2014): 1-11.

Diaz-Ordaz, K., Michael Kenward, and Richard Grieve. "Handling Missing Values in Cost Effectiveness Analyses That Use Data from Cluster Randomized Trials." *Journal of the Royal Statistical Society Series A* 177.Part 2 (2014): 457-74.

Diggle, P., and M.G. Kenward. "Informative Drop-Out in Longitudinal Data Analysis. " *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 43.1 (1994): 49-93.

Faria, Rita, Manuel Gomes, David Epstein, and Ian White. "A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials." *PharmacoEconomics* July 2014 (2014): n. pag. SpringerLink. Web. 21 Aug. 2014.

Flynn, Terry N., and Tim J. Peters. "Cluster randomized trials: Another problem for cost-effectiveness ratios." *International Journal of Technology Assessment in Health Care* 21.3 (2005): 403-409.

Gilleskie, Donna, and Thomas Mroz. "A flexible approach for estimating the effects of co-variates on health expenditures." *Journal of Health Economics* 23 (2004): 391-418.

Goldstein, Harvey. *Multilevel Statistical Models.* 2nd ed. Hoboken: John Wiley and Sons, 1995.

Gomes, Manuel, Edmond S.W. Ng, Richard Grieve, Richard Nixon, James Carpenter, and Simon G. Thompson. "Developing Appropriate Methods for Cost-Effectiveness Analysis of Cluster Randomization Trials." *Medical Decision Making* 32 (2011): 350-361.

Gomes, Manuel, Karla Diaz-Ordaz, Richard Grieve, and Michael Kenward. "Multiple Imputation Methods for Handling Missing Data in Cost-Effectiveness Analysis That Use Data from Hierarchical Studies: An Application to Cluster Randomized Trials." *Medical Decision Making* 33 (2013): 1051-063.

Heckman, James. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5.4 (1976): 475-93.

Heckman, James. "Sample Selection Bias as a Specification Error." *Econometrica* 47.1 (1979): 153-61.

Kessler, Ronald C., Catherine Barber, Phillip Wang, Arne Beck, Patricia Berglund, Paul Cleary, David McKenas, Nico Pronk, Gregory Simon, Paul Stang, and T. Bedirhan Ustun. "The World Health Organization Health and Work Performance Questionnaire (HPQ)." *Journal of Occupational and Environmental Medicine* 45.2 (2003): 156-174.

Kessler, Ronald, Minnie Ames, Pamela Hymel, Ronald Loeppke, David McKenas, Dennis Richling, Paul Stang, and T. Bedirhan Ustun . "Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to Evaluate the Indirect Workplace Costs of Illness." *Journal of Occupation and Environmental Medicine* 46.6 (2004): S23-S37.

Koopmanschap, Marc, and B. Martin van Ineveld. "Towards a New Approach for Estimating Indirect Costs of Disease." *Social Science and Medicine* 34.9 (1992): 1005-1010.

Koopmanschap, Marc, Frans Rutten, B. Martin van Ineveld, and Leona van Roijen. "The friction cost method for measuring indirect cost of disease." *Journal of Health Economics* 14 (1995): 171-189.

Little, Roderick, and Donald Rubin. *Statistical Analysis With Missing Data.* 2002 ed. Hoboken: John Wiley & Sons Inc, 2002.

Manning, Willard, and John Mullahy. "Estimating log models: to transform or not to transform?" *Journal of Health Economics* 20 (2001): 461-494.

Manning, Willard, Anirban Basu, and John Mullahy. "Generalized modeling approaches to risk adjustment of skewed outcomes data." *Journal of Health Economics* 24 (2005): 465-488.

Mihaylova, Borislava, Andrew Briggs, Anthony O'Hagan, and Simon Thompson. "Review of statistical methods for analysing health care resources and costs." *Health Economics* 20 (2011): 897-916.

Molenberghs, Geert. "Sensitivity Analysis of Continuous Incomplete Longitudinal Outcomes." *Statistica Neerlandica* 57.1 (2003): 112-35.

Molenberghs, Geert, and Michael Kenward. *Missing Data in Clinical Studies.* West Sussex: John Wiley and Sons, 2007.

Noble, Sian Marie, William Hollingworth, and Kate Tilling. "Missing data in trial-based cost-effectiveness analysis: the current state of play." *Health Economics* 21 (2012): 187-2012.

Robroek SJW, Lenthe FJV, Empelen PV, Burdorf A: Determinants of participation in worksite health promotion programmes: a systematic review. Int J *Behav Nutr Phy* 2009, 6:26.

Rubin , Donald. "Inference and missing data." *Biometrika* 63.3 (1976): 581-92.

Rubin , Donald. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91.434 (1996): 473-489.

Schafer, Joseph. "Multiple imputation: a primer ." *Statistical Methods in Medical Research* 8 (1999): 3-15.

Vella, Francis. "Estimating Models with Sample Selection Bias: A Survey." *The Journal of Human Resources* 33.1 (1998): 127-69.

Verbeke, Geert, Emmanuel Lesaffre, and Bart Spiessens. "The Practical Use of Different Strategies to Handle Dropout in Longitudinal Studies." *Drug Information Journal* 35 (2001): 419-34.

Williams, Geoffrey C., Virginia M. Grow, Zachary R. Freedman, Richard M. Ryan, and Edward Deci. "Motivational Predictors of Weight Loss and Weight-Loss Maintenance." *Journal of Personality and Social Psychology* 70.1 (1996): 115-26.

Wooldridge, Jeffery. *Econometric Analysis of Cross Section and Panel Data.* 1st ed. Cambridge: MIT, 2002.

You, Wen, Fabio Almeida, Jamie Zoellner, Jennie L. Hill, Courtney Pinard, Kacie Allen, Russell Glasgow, Laura Linnan, and Paul Estabrooks. "Who Participates in Internet-based Worksite Weight Loss Programs?" *BMC Public Health* 11.709 (2011): n. pag.