**Estimating causal effects in incomplete observational studies using multiple imputation and propensity score analysis: A simulation study**

Alessandra Valcarcel, University of Connecticut, Department of Statistics

May 14, 2015

**ACKNOWLEDGEMENTS**:

I would like to express my gratitude to my advisor, Dr. Ofer Harel, whose expertise, understanding, and patience, added considerably to my undergraduate research experience. Without his initial faith in taking me on as an undergraduate researcher, this thesis and my future graduate career would not be possible. I appreciate his vast knowledge and skill in many areas and his assistance in writing my thesis. I would like to thank the other members of the research group, Dr. Tania Huedo-Medina, Dr. Haim Bar, and Dr. Elizabeth Schifano for the assistance they provided at all levels of the research project. Finally, I would like to thank Johanna Claps for taking time out from her busy schedule to serve as my external reader and teaching me the value of editing.

I would also like to thank my family for the support they provided me throughout my entire life. In particular, I must acknowledge my parents without whose lessons in being feisty, independent, persevering, and editing assistance, I would not have finished this thesis. Finally, I would like to express my gratitude to my grandparents who taught me knowledge is power and inspiring my intellectual curiosity.

**Abstract:**

Incomplete data is a common problem in data analysis and can lead to complications in estimation, validity, and inference. I evaluate the joint use of two missing data methods, complete case analysis (CCA) and multiple imputation (MI) prior to propensity score analysis in incomplete observational data. Although there are many different approaches to dealing with incomplete data, complete case analysis is commonly used in medical studies and multiple imputation has become a leading method. Recently, propensity score analysis (PSA) is gaining popularity in inferring causality when an observational data was collected. Propensity score analysis must be done with a complete data. I assess the estimation, precision, and validity of results when incomplete data methods are used in conjunction with propensity score analysis. Additionally, I evaluate two methods of implementing multiple imputation, one method excludes the response variable from the imputation while the other method includes the response variable. These methods are chosen to investigate the impact of excluding the response in the imputation model.

**Introduction:**

Causal inference is the main goal of many medical and epidemiological investigations. Causal inference is a branch of statistics developed in order to identify treatments that cause a response or an outcome. When a causal relationship is identified, it leads to a better understanding of a response and thus intervention and prevention plans can be more effective. Causal inference answers questions such as: Does this treatment work? Is a certain exposure harmful? If yes, to what extent? In order to infer causality, certain types of clinical trials and associated analysis methods are necessary.

The gold standard approach to inferring causality is implementing a randomized controlled trial. A randomized controlled trial is a study design in which people are allocated to receive either treatment or control interventions at random (Stolberg et al. 2004). The ideal randomized controlled trial would include the following properties: treatment is allocated at random to subjects, all subjects are perfectly compliant, all relevant data are collected, and all relevant data are measured without error. The employment of a randomized controlled trial permits statisticians to discount “chance” alone as an explanation allowing the conclusion that any observed effects are causal on the response variable. Although randomized controlled trials seem simple, often times a randomized controlled trial is impractical.

In reality, randomized controlled trials are often ethically, practically, and/or economically infeasible. For example, in evaluating blood pressure in pregnant women, it is problematic to randomize women into pregnant and non-pregnant groups. In order to collect data in these situations, researchers select individuals with the treatment characteristic, for example pregnant women. After groups are selected, associated covariates and response variables of interest are measured for all subjects, and then the treatment and control are analyzed as distinct groups. This is an example of an observational study design.

Observational studies are a solution to many ethical, practical, and economical concerns associated with randomized controlled trials. According to the *Dictionary of Epidemiology* (2014), an observational study is a study in which assignment of treatment or control to subjects is outside the control of the investigator (Porta et al. 2014). Observational studies contrast a randomized controlled trial design where subjects are randomly allocated to groups. Groups selected to be in the treatment group based on a characteristic can lead to systematic differences among the groups, also known as selection bias. Often times, this bias will be interpreted as the effect of treatment, yet it is simply a bias in the sample selection due to non-randomized samples (Rosenbaum, 2002). Continuing with the pregnancy example, perhaps the difference in blood pressure is not due to pregnancy, but due to the fact that most pregnant women are in a similar age group. Observational studies must be manipulated to infer causality.

Methods have been developed to make causal inference on observational data. One such method is propensity score analysis (PSA). Propensity score analysis allows investigators to infer causality on observational data by attempting to reduce bias among groups. Treatment and control groups are balanced by matching similar subjects on observed covariates. Propensity score analysis calculates a specific “score” for each subject and then attempts to match individuals with similar scores in the treatment and control groups. Scores that cannot be matched are often discarded but can be included in analysis. Once data are matched, the covariates in the groups are balanced as if data were collected using a randomized controlled trial. When data meet the randomization assumption, then causal inference can be applied even in the absence of a randomized controlled trial.

A propensity score is a mathematical concept and thus has a formal definition. Mathematically, a propensity score is the conditional probability of treatment assignment given all observed covariates. Suppose treatment is binary where, 1 represents treatment, and 0 represents control, is a matrix of observed covariates, and finally, is the response variable. The propensity score calculation is described below:

.

The propensity score can be calculated by running a logistic regression where treatment is the binary outcome and all observed covariates are used in the model as predictors. Once the propensity score is calculated, the scores in the treatment group are matched with scores in the control group based on certain criteria of matching. Some common matching schemes include exact, nearest, nearest using a caliper, optimizing overall group difference, stratification matching, Mahalanobis metric matching, and full matching described extensively by Guo and Fraser (2009). If a propensity score is calculated with all known relevant covariates, and matched decisively, then the systematic bias in the two groups is eliminated. At this point, causality can be inferred with normal statistical analysis.

Propensity score analysis, like most methods in statistics, was developed to analyze rectangular or complete data (Little and Rubin, 1987). When data are incomplete, complications arise in the use of propensity score methods. One approach to deal with the incomplete data is complete case analysis, which deletes any subject with a missing value. Another approach involves complete case analysis then reweighting the remaining known subjects’ values. Alternatively, missing values can be estimated using the distribution of the complete data. Estimation can be done numerous ways such as imputation described by Nordholt (1998), maximum likelihood methods described by Dempster (1977), and averaging available items described by Schafer et al. (2002). The goal of multiple imputation (MI) is to generate valid inference in situations where data is incomplete (Harel et al. 2007).

Multiple imputation is one principled method for handling incomplete data. Multiple imputation is a simulation based approach, which entails creating several complete datasets by replacing missing values with a set of plausible estimates. The missing values are generated using a model, which is composed of the observed data, including all covariates and response variables. For example, data are collected for age and salary with the response body weight. The most inclusive model to use in multiple imputation includes age, salary, and body weight. The imputed values are then used in place of the missing values to generate complete datasets. The analysis is then carried out on each of the individual datasets and the results are to be combined using Rubin’s rules (Rubin, 1987).

In Rosenbaum and Rubin’s (1983) original proposal of propensity score analysis, they detailed the assumptions associated with implementing propensity score analysis. These assumptions need to be met in order for the analysis to be valid.

**Assumptions**:

1. Conditional on the covariates , the treatment variable is independent of the error term. This is called the ignorable treatment assignment assumption.
2. There is no correlation between the covariates and the error term .
3. There is no correlation between the covariates and treatment variable .
4. The error term is identically and independently distributed and follows a standard normal distribution (Guo and Fraser, 2009).

Multiple imputation also has associated assumptions that must be met. Most multiple imputation techniques assume data to be missing at random (MAR). Missing at random is defined as, given the observed data, the missingness mechanism does not depend on the unobserved data. Equivalently, this can be described as the behavior of two subjects who share observed variables have the same statistical behavior on other observations, whether variables are missing or not. Another assumption associated with multiple imputation is the model used in imputation must be “correct” in order to generate accurate estimates (Meng, 1994). Lastly, the model used in the imputation must match, in some sense, the model used in the analysis. Rubin (1987, 1996) describes these assumptions in depth.

Extensive research has been conducted on using propensity score analysis when data are complete such as Rosenbaum and Rubin (1983), Guo and Fraser (2009), and Wei and Haiyan (2015) but little has been investigated when data are incomplete. D’Agostino and Rubin (2000) use a maximum likelihood estimation approach to estimate missing data points prior to propensity score analysis. Mitra and Reiter (2011) estimate propensity score analysis with incomplete covariate data using general location mixture models but find under many missing data patterns there may not be sufficient data for accurate estimation. Mitra and Reiter (2012) compare the joint use of multiple imputation and propensity score analysis and investigated two methods of combing the results. However, Mitra and Reiter (2012) did not investigate the impact of including the response in the imputation model.

Literature concerning multiple imputation illustrates the importance of including response in the imputation model in order to meet the “correct” model assumption (Meng, 1994). In addition, multiple imputation assumes that the imputation model matches the model used in analysis (Rubin, 1987). However, propensity score analysis generates the conditional probability of treatment given all covariates excluding the response variable (Guo and Fraser, 2009). Thus, the assumptions of multiple imputation and propensity score analysis are contradicting each other. It is not clear if we need to follow multiple imputation assumptions and include the response variable in the imputation model, or follow the propensity score analysis assumptions and exclude the response variable.

This paper will address the implications of using multiple imputation prior to propensity score analysis. The questions that will be addressed are:

* What happens to bias and rates of missing information when multiple imputation excludes and includes the response variable?
* What happens to the validity of the results when treatment is dependent on a variable with missing information? Which missing data schemes will have valid results?

**Methods**:

In order to implement propensity score analysis, there are 3 major stages. First, propensity scores are calculated for each subject using an inverse logistic regression. Here the inverse logistic regression model used is:

where treatment is binary. This step calculates the probability of a subject being in the treatment group given the observed covariates. After propensity scores are calculated for each subject, subjects are matched based on certain matching schemes. The matching scheme carried out here is one-to-one nearest neighbor. One-to-one nearest neighbor matches one control subject to one treatment subject by finding a control subject with the closest propensity score. Other matching schemes were implemented such as optimal, exact, and full. These other methods produced similar results. All subjects not matched are discarded so sample sizes in the control and treatment group are equal. Once the treatment subjects are matched with the control subjects, two complete groups are ready for analysis and comparison. The calculations of the sample statistics are described in more detail in the Simulation Section and are used to analyze the data. The sample statistics calculated are mean difference between the treatment and control groups, pooled standard deviation, standard error of the estimated treatment effects, confidence intervals and coverage for the difference in the control and treatment groups. The data are generated so the true treatment effect .

Propensity score analysis requires a complete, or rectangular, dataset in order to be implemented. Thus, in the case of incomplete data complete case analysis is evaluated. In this case, any subject with a missing value is deleted. After the deletion of any subject with a missing value, the data are rectangular and propensity score analysis can be carried out as described previously (though the sample size changes). The calculations of the sample statistics are described in more detail in the Simulation Section and are used to analyze the data. The sample statistics calculated are mean difference between the treatment and control groups, pooled standard deviation, standard error of the estimated treatment effects, confidence intervals and coverage for the difference in the control and treatment groups. The data are generated so the true treatment effect .

In addition to complete case analysis, multiple imputation is evaluated prior to propensity score analysis to generate a complete data. The first step of multiple imputation is imputing missing values from a specified model containing the observed variables. The imputation model used is bayesian linear regression to generate missing values. Bayesian linear regression is used because data are normally distributed. Two imputation models are compared for each simulation setting; one contains the response variable while the second excludes the response variable . The bayesian linear regression approach imputes values based on the joint posterior distribution of the parameters and the missing data, which is conditional on the observed data. When the response variable is excluded from the model, the model is described by:

.

When the response variable is included in the imputation model, the model is described by:

.

Data are evaluated for 3 multiple imputation settings, , , in order to evaluate the use of a “small”, “moderate”, and “large” imputation number. After multiple imputation, datasets are independently complete and analysis can be carried out.

Once data are complete using multiple imputation, propensity score analysis is implemented. Propensity score analysis is performed on each data individually as described above. After propensity score analysis, matched data are obtained. The calculations of the sample statistics are described in more detail in the Simulation Section and are used to analyze the data. The sample statistics calculated are mean difference between the treatment and control groups, pooled standard deviation, standard error of the estimated treatment effects, confidence intervals and coverage for the difference in the control and treatment groups for each imputation. The data are generated so the true treatment effect .

Since the estimated treatment effect is the statistic of interest, the results from the analysis must be combined. Rubin’s rules for combining these multiply imputed estimates are based on asymptotic theory. Using Rubin’s original notation, for a single population parameter of interest , in our case treatment effect, the multiple imputation point estimate is the average of the estimates of from the imputed datasets. Let and represent the mean difference in treatment and control groups, and the estimated variance for each of the datasets. The combined mean difference in treatment and control groups, or point estimate is described below by the formula:

.

The associated total variance is denoted by and calculated by the formula:

,

where

is the within imputation variance and

is the between imputation variance. After the mean difference between treatment and control and pooled standard deviations are combined using Rubin’s rules, the confidence interval, and coverage can be calculated. Additionally, when multiple imputation is implemented, the rate of missing information is calculated using Rubin’s rules (Rubin, 1987). At this point, estimates for the mean difference and pooled standard deviation are combined. Once a single estimate is calculated for the mean difference and pooled standard deviation, the standard error of the estimated treatment effects, confidence intervals and coverage for the difference in the control and treatment groups can be calculated. The calculations of the sample statistics are described in more detail in the Simulation Section and are used to analyze the data. The data are generated so the true treatment effect .

**Simulation Study:**

The statistical software known as R is utilized to generate artificial data so parameters and distributions are known (R Core Team, 2014). For each simulation run, two covariates, and are generated with . Each covariate, and , is generated from a normal distribution with mean ), variance , and correlation (0.5). Data is then replicated independently times in order to evaluate validity in repeated trials.

Treatment is assigned in order to generate exactly 100 treated subjects and 1000 control subjects.

The response variable is calculated in the control group using the model:

,

where ,.

Thus, the expected value of the response variable in the control group is .

The response variable is calculated in the treatment group using the model:

,

where ,, and is the true treatment effect.

Three mechanisms are considered to generate the treatment effect :

1. Treatment effect depends only on given by the model:

(1)

where and .

1. Treatment effect depends only on given by the model:

(2)

where and .

1. Treatment effect depends equally on and given by the model:

(3)

where , , and .

Treatment effect, or the difference in means of the treatment group and control group, is calculated in this way so that in all settings.

Missing data are introduced to based on missing at random mechanisms while and are fully observed. Within each treatment assignment setting, two mechanisms for assigning missing values in are considered:

1. Missing values are introduced only to control subjects from the Bernoulli distribution

.

1. Missing values are introduced only to treatment subjects from the Bernoulli distribution

.

In this way, subjects with larger values, which are the subjects most likely to be selected as matches, are more likely to be missing values. Missing data are introduced to 30% of control subjects. Similarly, missing data are introduced to 30% of treatment subjects.

The covariate , treatment variable , and response variable are assumed to be fully observed. Thus, two different imputation models are evaluated. The first model includes and while the second includes , , and . Excluding and including the response variable in the imputation model is performed in order investigate the impacts of excluding the response variable in the imputation. Multiple imputation is carried out and the complete datasets over replications.

Propensity score analysis is carried out within each replication ) independently using the R package ‘MatchIt’ (Ho et al. 2009). Propensity scores are calculated for each subject using a one-to-one nearest neighbor matching scheme.

**Calculation of Sample Statistics:**

The sample statistics reported are:

1. Point estimate (Pt. Est.) calculated by the formula:
2. Pooled standard deviation (Pooled Std.) calculated by the formula:

where and are the sample variances of the treatment and control group.

1. Standard Error of () is calculated by the formula:

where sd is the standard deviation.

1. 95% Lower Bound Confidence Interval (L.B.) is calculated by the formula:
2. 95% Upper Bound Confidence Interval (U.B.) is calculated by the formula:
3. Length is calculated by the formula:
4. Coverage (Cov.) is calculated by summing when the true treatment effect ( is in the confidence interval within each replicated simulation.
5. Rate of Missing Information (Rt. M. I.) is calculated from the R function ‘pool.scalar’ in the package ‘mice’ (Buuren and Groothuis-Oudshoorn, 2011). Rubin’s rules are used to combine results and then sample statistics are then calculated.

**Results:**

Results are reported for simulation (I) (based on equation (1)). In simulation (I) treatment assignment is dependent on . Missingness is then introduced on control units and then separately on treatment units . Settings are run with the response excluded and included from the imputation model. Results for simulation (II) and (III) (based on equations (2) and (3)) are reported in the Appendix.

**Simulation (I): Treatment assignment depends only on**

1. **Missingness is introduced on the variable in the control group only; true treatment effect**

**Table 1.1a:** Missingness is introduced on the  variable in the control group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.506 | 0.301 | 0.368 | 4.916 | 6.097 | 1.181 | 0.569 | NA |
| **MI** |  | 5.024 | 0.406 | 0.147 | 4.227 | 5.820 | 1.593 | 1 | 0.376 |
|  | 5.021 | 0.401 | 0.112 | 4.237 | 5.806 | 1.569 | 1 | 0.358 |
|  | 5.024 | 0.398 | 0.107 | 4.248 | 5.806 | 1.558 | 1 | 0.352 |

The point estimates in this case are unbiased except for complete case analysis, which is inflated. The pooled standard deviation is valid for complete case analysis but slightly higher for all the multiple imputation settings. Standard error of the treatment effect is lower in the multiple imputation setting than the complete data and complete case analysis. This implies that when data is multiple imputed, it is more stable. The length of the confidence intervals for the complete data and complete case analysis are similar. In multiple imputation, the length of the confidence interval is slightly wider so coverage is 100% but this is due to the wider confidence interval. Coverage for multiple imputation is comparable to the complete data but coverage for complete case analysis is poor.

**Table 1.1b:** Missingness is introduced on the  variable in the control group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.506 | 0.301 | 0.368 | 4.916 | 6.097 | 1.181 | 0.569 | NA |
| **MI** |  | 5.020 | 0.379 | 0.118 | 4.276 | 5.762 | 1.486 | 1 | 0.270 |
|  | 5.023 | 0.376 | 0.089 | 4.286 | 5.761 | 1.475 | 1 | 0.259 |
|  | 5.021 | 0.375 | 0.084 | 4.285 | 5.756 | 1.471 | 1 | 0.256 |

The point estimates in this case are unbiased except for complete case analysis, which is inflated and biased. The pooled standard deviation is valid for complete case analysis but slightly higher for all the multiple imputation settings. Standard error of the treatment effect is lower in the multiple imputation setting than the complete data and complete case analysis. This implies that when data is multiple imputed, it is more stable. The length of the confidence intervals for the complete data and complete case analysis are similar. In multiple imputation, the length of the confidence interval is slightly wider so coverage is 100% but this is due to the wider confidence interval. Coverage for multiple imputation is comparable to the complete data but coverage for complete case analysis is poor. Comparing the rate of missing information when response is excluded or included in the imputation model, the rates of missing information are lower when response is included.

1. **Missingness is introduced on the variable in the treatment group only; true treatment effect**

**Table 1.2a:** Missingness is introduced on the  variable in the treatment group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.015 | 0.328 | 0.267 | 4.372 | 6.658 | 2.286 | 0.981 | NA |
| **MI** |  | 4.976 | 0.414 | 0.180 | 4.163 | 5.799 | 1.636 | 0.999 | 0.393 |
|  | 5.035 | 0.410 | 0.159 | 4.230 | 5.840 | 1.610 | 1 | 0.387 |
|  | 5.018 | 0.408 | 0.156 | 4.216 | 5.820 | 1.604 | 1 | 0.382 |

The point estimates in this case are unbiased. The pooled standard deviation is valid for complete case analysis but slightly higher for all the multiple imputation settings. Standard error of the treatment effect is lower in the multiple imputation setting than the complete data and complete case analysis. This implies that when data is multiple imputed, it is less variable. The length of the confidence interval for complete case analysis is wider than when control units were missing explaining the higher coverage. Coverages for the complete data, complete case analysis, and multiple imputation are comparable.

**Table 1.2b:** Missingness is introduced on the  variable in the treatment group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.015 | 0.328 | 0.267 | 4.372 | 6.658 | 2.286 | 0.981 | NA |
| **MI** |  | 5.014 | 0.378 | 0.112 | 4.272 | 5.756 | 1.484 | 1 | 0.268 |
|  | 5.020 | 0.376 | 0.094 | 4.282 | 5.757 | 1.475 | 1 | 0.256 |
|  | 5.019 | 0.374 | 0.088 | 4.283 | 5.753 | 1.470 | 1 | 0.251 |

The point estimates in this case are unbiased. The pooled standard deviation is valid for complete case analysis but slightly higher for all the multiple imputation settings. Standard error of the treatment effect is lower in the multiple imputation setting than the complete data and complete case analysis. This implies that when data is multiple imputed, it is more stable. The length of the confidence interval for complete case analysis is wider than when control units were missing explaining the higher coverage. Coverages for the complete data, complete case analysis, and multiple imputation are comparable. Comparing the rate of missing information when response is included or excluded, it is higher when response is excluded.

**Conclusion:**

Randomized controlled trials are the gold standard approach to causal inference. Often times, randomized controlled trials cannot be implemented due to practical, ethical, and/or economical limitations. When randomized controlled trials are not possible, researchers turn to observational studies. Causal inference can be estimated on observational studies with the use of propensity score analysis. Propensity score analysis attempts to reduce bias due to confounding variables. Once confounding variables are neutralized, causal inference can be made.

Propensity score analysis requires a complete data set yet; missing data is a common complication is data collection and analysis. Multiple imputation is a simulation based approach to deal with incomplete data. Missing values are estimated using a model, which is ideally composed of all covariates and response variables. independent complete datasets are generated. From these complete datasets, analysis is carried out on each and results are to be combined using Rubin’s rules (Rubin, 1987).

The combined use of propensity score analysis and multiple imputation poses some clashing assumptions. Literature concerning multiple imputation illustrates the importance of including response in the imputation model in order to meet the “correct” model assumption (Meng, 1994). However, propensity score analysis generates the conditional probability of treatment given all covariates excluding the response variable (Guo and Fraser, 2009). Thus, the assumptions of multiple imputation and propensity score analysis are contradicting each other.

In the joint use of multiple imputation and propensity score analysis researchers must address the issues of including response in the multiple imputation model. Including response in the imputation model meets the assumptions in multiple imputation but then violates assumptions in propensity score analysis. When response is included in the imputation model, the treatment effect and rates of missing information change but the validity of results must be checked.

The joint use of multiple imputation and propensity score analysis will generate unbiased point estimates and a reliable standard deviation. Complete case analysis often had biased point estimates and inflated standard deviation and is not recommended under these conditions. While multiple imputation generates seemingly reliable sample statistics, coverage is considered too high due to the confidence intervals being too wide. The wider confidence intervals lead to coverage of 100% in all the multiple imputation settings when they should be at 95%. While 100% coverage is better than low coverage, the confidence intervals should be narrower and have a coverage of 95%. Using a “small”, “moderate”, and “large” imputation model generates essentially the same results but was slightly more accurate in all settings because the point estimate most accurately estimate the true treatment effect, pooled standard deviation was smallest, standard error of the treatment effect was the lowest, and the rate of missing information was smallest.

Including the response variable in the multiple imputation model generated more accurate point estimates with less variation. This can be concluded because the standard error of the mean point estimates is lower when response was included in the imputation model. Additionally, the pooled standard deviation was slightly smaller when response was included in the multiple imputation model thus the data was more consistent. The rate of missing information loss was also lower when response was included in the model. When response is excluded from the imputation model, rate of missing information is about 36% while when response is included the rate drops to 26% indicating less information is lost when response is included in the imputation model. Under the conditions of this study, including the response variable in the imputation model led to unbiased results and analysis and outperformed the exclusion of response in the imputation model.

Using multiple imputation in conjunction with propensity score analysis does have some limitations. With coverage at 100%, this indicates the confidence intervals obtained after multiple imputation are too wide. Coverage is expected to be and inflated coverage indicates how well point estimates and standard deviations are estimated.

While including the response in the imputation model generates accurate sample statistics with less variation, it still has drawbacks. The limitation of including response in the imputation model followed by propensity score analysis is it still violates the assumptions of propensity score analysis. Propensity score analysis generates the conditional probability of treatment given all covariates excluding the response variable (Guo and Fraser, 2009). The investigator must then chose whether they prefer to include response in the imputation model and violate the assumptions of propensity score analysis or exclude response and have slightly less accurate results.

**Future work:**

Currently, data is missing at a rate of 30% of either the control or treatment units. Future work can investigate the consequences of increases the rate of missing data. In this paper, sample size is 1100 and the treatment unit to control unit ratio is 1:10. With this ratio, matching balances the treatment and control groups well because there are more subjects pick as a match in the control group. Varying the ratio of treatment unit to control unit can lead to different results. For example, if the ratio is 1:1 then matching after propensity score analysis may not lead to balanced groups. Also, varying the treatment effect to be smaller may be of interest. When there is little difference between groups, the multiple imputation prior to propensity score analysis may not be sensitive to the small treatment effect and falsely show no difference.

**References:**

Buuren, S., & Groothuis-Oudshoorn, K. (2011). *MICE: Multivariate imputation by chained equations in R* American Statistical Association. Retrieved from [http://cran.r-project.org/web/packages/mice/mice.pdf](http://cran.r-project.org/web/packages/mice/mice.pdf" \t "_blank)

D'Agostino Jr, R. B., & Rubin, D. B. (2000). Estimating and using propensity scores with partially missing data. *Journal of the American Statistical Association, 95*(451), 749-759.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society.Series B (Methodological),* , 1-38.

Guo, S., & Fraser, M. W. (2009). *Propensity score analysis: Statistical methods and applications* Sage Publications.

Harel, O., & Zhou, X. (2007). Multiple imputation: Review of theory, implementation and software. *Statistics in Medicine, 26*(16), 3057-3077.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2013). *MatchIt: Nonparametric preprocessing for parametric causal inference (version 2.211)[software]* (2.4-21 ed.) Retrieved from [http://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf](http://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf" \t "_blank)

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data* John Wiley & Sons.

M. Davidian. (2005). Simulation studies in statistics. Retrieved from [http://www4.stat.ncsu.edu/~davidian/st810a/simulation\_handout.pdf](http://www4.stat.ncsu.edu/~davidian/st810a/simulation_handout.pdf" \t "_blank)

Meng, X. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science,* , 538-558.

Mitra, R., & Reiter, J. P. (2011). Estimating propensity scores with missing covariate data using general location mixture models. *Statistics in Medicine, 30*(6), 627-641.

Mitra, R., & Reiter, J. P. (2012). A comparison of two methods of estimating propensity scores after multiple imputation. *Statistical Methods in Medical Research,* doi:0962280212445945 [pii]

Nordholt, E. S. (1998). Imputation: Methods, simulation experiments and practical examples. *International Statistical Review, 66*(2), 157-180.

Pan, W., & Bai, H. (2015). *Propensity score analysis: Fundamentals and developments* Guilford Publications.

Porta, M., Greenland, S., Hernán, M., Silva, I. D. S., & Last, J. M. (2014). *A dictionary of epidemiology* Oxford University Press.

R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing,*

Rosenbaum, P. R. (2002). *Observational studies* Springer.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* John Wiley & Sons.

Rubin, D. B. (1996). Multiple imputation after 18 years. *Journal of the American Statistical Association, 91*(434), 473-489.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147.

Stolberg, H. O., Norman, G., & Trop, I. (2004). Randomized controlled trials. *American Journal of Roentgenology, 183*(6), 1539-1544.

**Appendix I: Results for All Simulation Settings  
Simulation (I): Treatment assignment depends only on**

1. **Missingness is introduced on the variable in the control group only; true treatment effect**

**Table 1.1a:** Missingness is introduced on the  variable in the control group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.506 | 0.301 | 0.368 | 4.916 | 6.097 | 1.181 | 0.569 | NA |
| **MI** |  | 5.024 | 0.406 | 0.147 | 4.227 | 5.820 | 1.593 | 1 | 0.376 |
|  | 5.021 | 0.401 | 0.112 | 4.237 | 5.806 | 1.569 | 1 | 0.358 |
|  | 5.024 | 0.398 | 0.107 | 4.248 | 5.806 | 1.558 | 1 | 0.352 |

**Table 1.1b:** Missingness is introduced on the  variable in the control group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.506 | 0.301 | 0.368 | 4.916 | 6.097 | 1.181 | 0.569 | NA |
| **MI** |  | 5.020 | 0.379 | 0.118 | 4.276 | 5.762 | 1.486 | 1 | 0.270 |
|  | 5.023 | 0.376 | 0.089 | 4.286 | 5.761 | 1.475 | 1 | 0.259 |
|  | 5.021 | 0.375 | 0.084 | 4.285 | 5.756 | 1.471 | 1 | 0.256 |

1. **Missingness is introduced on the variable in the treatment group only; true treatment effect**

**Table 1.2a:** Missingness is introduced on the  variable in the treatment group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.015 | 0.328 | 0.267 | 4.372 | 6.658 | 2.286 | 0.981 | NA |
| **MI** |  | 4.976 | 0.414 | 0.180 | 4.163 | 5.799 | 1.636 | 0.999 | 0.393 |
|  | 5.035 | 0.410 | 0.159 | 4.230 | 5.840 | 1.610 | 1 | 0.387 |
|  | 5.018 | 0.408 | 0.156 | 4.216 | 5.820 | 1.604 | 1 | 0.382 |

**Table 1.2b:** Missingness is introduced on the  variable in the treatment group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.001 | 0.318 | 0.214 | 4.385 | 5.634 | 1.249 | 0.990 | NA |
| **CCA** | | 5.015 | 0.328 | 0.267 | 4.372 | 6.658 | 2.286 | 0.981 | NA |
| **MI** |  | 5.014 | 0.378 | 0.112 | 4.272 | 5.756 | 1.484 | 1 | 0.268 |
|  | 5.020 | 0.376 | 0.094 | 4.282 | 5.757 | 1.475 | 1 | 0.256 |
|  | 5.019 | 0.374 | 0.088 | 4.283 | 5.753 | 1.470 | 1 | 0.251 |

**Simulation (II): Treatment assignment depends only on**

1. **Missingness is introduced on the variable in the control group only; true treatment effect**

**Table 2.1a:** Missingness is introduced on the  variable in the control group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.000 | 0.319 | 0.214 | 4.374 | 5.625 | 1.251 | 0.989 | NA |
| **CCA** | | 5.496 | 0.301 | 0.368 | 4.906 | 6.088 | 1.182 | 0.571 | NA |
| **MI** |  | 5.014 | 0.406 | 0.147 | 4.217 | 5.811 | 1.594 | 1 | 0.375 |
|  | 5.012 | 0.400 | 0.112 | 4.226 | 5.797 | 1.571 | 1 | 0.357 |
|  | 5.015 | 0.399 | 0.107 | 4.233 | 5.797 | 1.564 | 1 | 0.351 |

**Table 2.1b:** Missingness is introduced on the  variable in the control group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.000 | 0.319 | 0.214 | 4.374 | 5.625 | 1.251 | 0.989 | NA |
| **CCA** | | 5.496 | 0.301 | 0.368 | 4.906 | 6.088 | 1.182 | 0.571 | NA |
| **MI** |  | 5.013 | 0.381 | 0.114 | 4.266 | 5.760 | 1.494 | 1 | 0.278 |
|  | 5.009 | 0.377 | 0.089 | 4.270 | 5.750 | 1.300 | 1 | 0.262 |
|  | 5.010 | 0.376 | 0.081 | 4.273 | 5.747 | 1.474 | 1 | 0.256 |

1. **Missingness is introduced on the variable in the treatment group only; true treatment effect**

**Table 2.2a:** Missingness is introduced on the  variable in the treatment group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.000 | 0.319 | 0.214 | 4.375 | 5.625 | 1.250 | 0.989 | NA |
| **CCA** | | 5.009 | 0.329 | 0.260 | 4.365 | 6.654 | 2.289 | 0.981 | NA |
| **MI** |  | 4.967 | 0.415 | 0.180 | 4.153 | 5.781 | 1.628 | 0.999 | 0.393 |
|  | 5.026 | 0.411 | 0.159 | 4.220 | 5.831 | 1.611 | 1 | 0.387 |
|  | 5.009 | 0.409 | 0.156 | 4.206 | 5.811 | 1.605 | 1 | 0.382 |

**Table 2.2b:** Missingness is introduced on the  variable in the treatment group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.000 | 0.319 | 0.214 | 4.375 | 5.625 | 1.250 | 0.989 | NA |
| **CCA** | | 5.009 | 0.329 | 0.260 | 4.365 | 6.654 | 2.289 | 0.981 | NA |
| **MI** |  | 5.011 | 0.384 | 0.126 | 4.258 | 5.764 | 1.506 | 1 | 0.268 |
|  | 5.015 | 0.376 | 0.097 | 4.277 | 5.754 | 1.477 | 1 | 0.256 |
|  | 5.015 | 0.376 | 0.090 | 4.277 | 5.753 | 1.476 | 1 | 0.251 |

**Simulation (III): Treatment assignment depends equally on and**

1. **Missingness is introduced on the variable in the control group only; true treatment effect**

**Table 3.1a:** Missingness is introduced on the  variable in the control group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.005 | 0.317 | 0.214 | 4.383 | 5.627 | 1.244 | 0.989 | NA |
| **CCA** | | 5.490 | 0.300 | 0.366 | 4.903 | 6.079 | 1.176 | 0.574 | NA |
| **MI** |  | 5.020 | 0.403 | 0.145 | 4.228 | 5.812 | 1.584 | 1 | 0.377 |
|  | 5.019 | 0.398 | 0.112 | 4.238 | 5.800 | 1.562 | 1 | 0.357 |
|  | 5.018 | 0.397 | 0.104 | 4.240 | 5.800 | 1.560 | 1 | 0.351 |

**Table 3.1b:** Missingness is introduced on the  variable in the control group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.005 | 0.317 | 0.214 | 4.383 | 5.627 | 1.244 | 0.989 | NA |
| **CCA** | | 5.490 | 0.300 | 0.366 | 4.903 | 6.079 | 1.176 | 0.574 | NA |
| **MI** |  | 5.013 | 0.376 | 0.116 | 4.274 | 5.752 | 1.478 | 1 | 0.268 |
|  | 5.017 | 0.372 | 0.096 | 4.286 | 5.748 | 1.462 | 1 | 0.253 |
|  | 5.015 | 0.371 | 0.092 | 4.288 | 5.743 | 1.455 | 1 | 0.247 |

1. **Missingness is introduced on the variable in the treatment group only; true treatment effect**

**Table 3.2a:** Missingness is introduced on the  variable in the treatment group only and the response variable is excluded from the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.004 | 0.317 | 0.214 | 4.383 | 5.627 | 1.244 | 0.989 | NA |
| **CCA** | | 5.015 | 0.330 | 0.258 | 4.369 | 6.662 | 2.293 | 0.983 | NA |
| **MI** |  | 4.973 | 0.409 | 0.186 | 4.170 | 5.777 | 1.607 | 1 | 0.386 |
|  | 5.030 | 0.413 | 0.170 | 4.220 | 5.841 | 1.621 | 1 | 0.400 |
|  | 5.016 | 0.410 | 0.160 | 4.212 | 5.820 | 1.608 | 1 | 0.391 |

**Table 3.2b:** Missingness is introduced on the  variable in the control group only and the response variable is included in the imputation model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Pt. Est.** | **Pooled Std.** |  | **L.B.** | **U.B.** | **Length** | **Cov.** | **Rt. M. I.** |
| **Comp. Data** | | 5.004 | 0.317 | 0.214 | 4.383 | 5.627 | 1.244 | 0.989 | NA |
| **CCA** | | 5.015 | 0.330 | 0.258 | 4.369 | 6.662 | 2.293 | 0.983 | NA |
| **MI** |  | 5.008 | 0.377 | 0.127 | 4.269 | 5.747 | 1.478 | 1 | 0.267 |
|  | 5.015 | 0.373 | 0.097 | 4.283 | 5.754 | 1.471 | 1 | 0.252 |
|  | 5.014 | 0.372 | 0.087 | 4.285 | 5.743 | 1.458 | 1 | 0.247 |