BIAS REDUCTION USING PROPENSITY SCORE MATCHING IN OBSERVATIONAL DATA

BY
SAFIYA SADA SANI
M.SC/SCIEN/05323/2009-2010

BEING A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL,

AHMADU BELLO UNIVERSITY,

ZARIA, NIGERIA

IN PARTIAL FULFILMENT FOR THE AWARD OF MASTERS OF SCIENCE IN STATISTICS, DEPARTMENT OF MATHEMATICS, AHMADU BELLO UNIVERSITY, ZARIA

AUGUST, 2014

DEDICATION

This work is dedicated to my parents, Alhaji and Bilkisu Sada Sani.

DECLARATION

I hereby declare that the work in this thesis entitled "Bias Reduction using Propensity Score Matching in Observational Data" has been carried out by me, under the supervision of Prof. O. E. Asiribo and Dr. H. G. Dikko. All information derived from the literature has been duly acknowledged in the text and a list of references provided. To the best of my knowledge, no part of this thesis has been previously presented for another degree or diploma at any university.

Safiya Sada Sani	_____	_____
Student	(Signature)	(Date)

CERTIFICATION

This research thesis entitled "**Bias reduction using propensity score matching in observational data**" by **Safiya Sada Sani** meets the regulations governing the award of degree of Master of Science (M.Sc) of Ahmadu Bello University, Zaria, and is approved for its contributions to knowledge and literary presentation.

……	…….	……..
Chairman, Supervisory Committee Prof. O.E. Asiribo	Signature	Date
……	…….	……..
Member, Supervisory Committee Dr. H. G. Dikko	Signature	Date
……	…….	……..
External Examiner Dr. O. E. Olubusoye	Signature	Date
……	…….	……..
Head of Department Dr. B. Sani	Signature	Date
……	…….	……..
Dean, Postgraduate School Prof. A. A. Joshua	Signature	Date

ACKNOWLEDGMENTS

First and foremost, I wish to thank Allah (SWT) for everything. He and He alone made everything possible.

This section allows me the opportunity to voice my appreciation to those who have helped me in attaining my educational goals. Direct thanks go to my Supervisory Committee, headed by Prof. O. E. Asiribo, for their help and guidance. I feel fortunate to have had the opportunity to work with them. To the Prof., my gratitude cannot be quantified for he had been very patient and understanding. Dr H. G. Dikko, also a member of the committee, had in so many ways aided in my research. I shall forever remain grateful for the continuous support and encouragement I get from you Sir.

Special thanks to my husband, Sabo A. Abubakar, for his emotional and financial support. Prof. S. U. Gulumbe also deserves to be acknowledged for his own contribution as my lecturer and for his advice concerning the project. I wish to thank Dr A. Mohammed and Dr A. Yahaya for their criticisms and constructive comments. To John Onwuche, who also made his own contributions, my gratitude has no bounds. I would also seize this opportunity to express my sincere gratitude to Dr O. E. Olubusoye of the University of Ibadan (my assigned external examiner), without whose endorsement I would not have scaled through the final stage of the programme. He was most helpful. I shall not forget my colleagues at the Data Processing Unit in the Institute for Agricultural Research (IAR), Ahmadu Bello University for their continuous support and understanding while I took time to work on this project. May God Almighty reward you all.

ABSTRACT

In observational studies, "case-control groups" often exhibit imbalance on covariates. This covariate imbalance is confounded with treatments. It is difficult to attribute differences in responses to the "treatment" because the covariates are also believed to influence the response. Propensity score matching attempts to reduce the confounding effects of covariates, and so allows differences of responses to be attributed to differences of treatments. In addition, the values of the propensity scores can serve as a diagnostic tool to evaluate the comparability of the groups in a quantitative way. When two groups are being compared, the propensity score can be calculated as the predicted probability of group membership from a logistic regression. It represents the 'tendency' for an observation to be in one group or the other. By adjusting for the value of the propensity score in a linear model, one effectively adjusts for any group differences attributed to the variables used to create the propensity score. Here we present an experiment where propensity scores were used to adjust for differences between a case and a control group (treatment group and a non-randomized control group). Propensity scores were created using SPSS Version 16 Binary Logistic Regression Procedure on a Windows Vista platform. A linear model was also estimated using the same. Groups were compared using independent samples t-tests and chi-square tests as appropriate. Standardized differences were calculated and matching was done with Microsoft Excel Version 2007 on a Windows Vista platform. The results showed that the Propensity Score Matching was successful in reducing the bias on the covariates.

TABLE OF CONTENT

DEDICATION ..... **I**

DECLARATION ..... **II**

CERTIFICATION ..... **III**

ACKNOWLEDMENTS ..... **IV**

ABSTRACT ..... **V**

LIST OF TABLES ..... **IX**

LIST OF APPENDICES ..... **X**

CHAPTER 1 Introduction **1**

1.1 Statement of Problem 3

1.2 Aim and Objectives of the Study 4

1.3 Justification for the Study 5

1.4 Significance of the Study 5

1.5 Definition of Terms 5

CHAPTER 2 Literature Review **8**

2.1 Observational Studies 8

2.2 Case-Control Studies 9

2.2.1 Bias in Observational Case-Control Studies 9

2.2.2 Traditional Methods for Bias Reduction 10

2.3 Propensity Scoring 12

2.3.1 Propensity Score Definition 12

2.4 The Use of Propensity Score Technique 15

2.4.1 Matching with Propensity Scores 15

2.4.2 Stratification (Subclassification) with Propensity Scores 16

2.4.3 Regression (Covariance) Adjustments with Propensity Scores 19

2.5 Logistic Regression 21

2.5.1 Introduction 21

2.5.2 The Model 22

2.5.3 Uses of Logistic Regression 23

2.6 Preterm Birth 24

2.6.1 Consequences of Preterm Births 26
CHAPTER 3 Methodology 28
3.0 Introduction 28
3.1 Preliminary Statistical Analysis 31
3.1.1 Correlation Coefficients 31
3.2 Propensity Scores 33
3.2.1 Propensity Score Variable Identification 33
3.2.2 Variable Selection for Propensity Score Model 34
3.2.3 Propensity Score Modelling 35
3.3 Logistic Regression and Propensity Score 36
3.3.1 Elements of the Propensity Score 36
3.4 Study Population 38
3.5 Analysis Selection 39
3.5.1 Matching 40
3.5.2 Matching Metric 41
3.5.3 Matching Variable 41
3.5.4 Matching Algorithm 41
3.5.5 Matching Structure 42
3.5.6 Replacement of Comparison Subjects 43
3.5.7 Nearest Neighbour Matching on Estimated Propensity Score 43
3.6 Assessing the Matched Data 44
3.6.1 The Standardized Bias 44
3.6.2 Bias Reduction 45
3.6.3 Test on Means and Proportions 46
3.7 Analysis 46
3.7.1 Software 46
CHAPTER 4 Data Analysis and Interpretation 47
4.1 Correlations with Baby Status 47
4.1.1 Correlation of Status with Covariates 47
4.2 Logistic Regression on Baby's Status 48
4.3 Group Comparisons 51
CHAPTER 5 Summary, Conclusions and Recommendations 55
5.1 Summary 55

5.2 Conclusions....56
5.3 Recommendations....57
5.4 Contribution to Knowledge....57
5.5 Suggested Areas for Further Research....58
REFERENCES....59
Appendix I: List of Potential Matches for the 13 Cases....65
Appendix II: Figures of Propensity and Logit Distributions....67
Appendix III: Dataset used for the Study....71

LIST OF TABLES

Table 4.1: Correlation of Status with the Covariates 47
Table 4.2a: Logistic Regression on Baby's Status.....48
Table 4.2b: Model Summary and Goodness-of-fit Tests 49
Table 4.2c: Classification Table.....50
Table 4.3: Group Comparisons before Matching.....51
Table 4.4: Propensity Score Matching 52
Table 4.5: Group Comparisons after Matching.....53
Table 4.6: Reduction in Bias 54

LIST OF APPENDICES

Appendix I: List of Potential Matches for the 13 Cases 65

Appendix II: Figures of Propensity and Logit Distributions 67

Figure 1: The Distribution of Propensity Scores before Matching……...67

Figure 2: The Distribution of the Logits of Propensity Scores before Matching……..68

Figure 3: The Distribution of Propensity Scores after Matching……..69

Figure 4: The Distribution of the Logits of Propensity Scores after Matching…….70

CHAPTER 1

INTRODUCTION

In order to make group comparisons, the generally accepted pattern in research consists of the following method:

- Formation of treatment and experimental groups, sometimes with a single group serving as its own control.
- Mapping treatments to the groups.
- Analysing group differences.
- Generalising findings based on groups to tendencies among future individuals.

Defining groups is a crucial first step and once they are defined, one would want their composition to be identical. Statistical adjustments, often in the form of blocking variables, variables or covariate analysis could be used to adjust for the pre-treatment group differences.

The random assignment of treatment to groups before comparison is often resorted to because, in theory, this assures that the groups are identical. This, however, is not always practical and does not necessarily result in groups that are equivalent in terms of all the important covariates. It is the expected values of the covariates over numerous replications that are equal.

A substitute to random assignment is a matched-pairs design whereby each member of the first group is matched with a member of the second group on all the factors the researcher considers to be viable and important. According to Rudner & Peyton (2006), in a well-matched pair, it is as if we are using the same individual twice. When matching is adequate, the variables used for matching that might cause confounding

problems are controlled. Matching on many covariates is however difficult, especially if one is trying to obtain an exact match when some of the covariates are continuous.

The assignment of treatments to experimental subjects in observational studies is unselective and not arbitrary; which is why observational studies frequently provide biased estimation of treatment effects and have imbalance on covariates. In other words, treatment groups are often compared to non-randomized control groups whereby inference of causality in such studies can be biased. One can expect that treatment effects are confounded with factors that determine selection into treatment (case) and control groups.

In this work, the propensity score matching procedure (PSM) is used. This procedure allows researchers to pair treatment and control individuals based on their propensity scores. It is particularly useful for ‘cases of causal inference and sample selection bias in non-experimental settings in which: (i) few units in the non-experimental comparison group are comparable to the treatment units; and (ii) selecting a subset of comparison units similar to the treatment unit is difficult because units must be compared across a high dimensional set of pre-treatment characteristics.

Propensity score matching is a refined approach to a matched-pairs design (Rosenbaum & Rubin, 1985b; Rubin, 1997; Joffe & Rosenbaum, 1999). Covariates are pooled to produce a propensity score, and individuals in the treatment group are matched to individuals in the control group based on their propensity score. By using this method, one is weighting the variables by their relative significance and matching based on best possible combination, preferably than by similarly weighted individual variables. Rubin

(1997) has shown that when one matches on the composite propensity score, the group means and standard deviations on the covariates will also be equivalent.

1.1 STATEMENT OF THE PROBLEM

In observational studies, investigators have no control over the treatment assignment. The treated (case) and non-treated (that is, control) groups may have large differences on their observed covariates, and these differences can lead to biased estimates of treatment effects (D'Agostino, 1998).

This difficulty may to some extent be avoided if information on measured covariates is included in the design of a study (e.g. through matched sampling) or into estimation of the treatment effect (e.g. through stratification or covariance adjustment). These long-established methods of adjustment (matching, stratification and covariance analysis adjustment) may be insufficient in bias elimination as they can only use a limited number of covariates for adjustment. However, propensity scores, which provide a scalar summary of the covariate information, do not have this imperfection. In the statistical analysis of observational data, propensity score analysis is a methodology attempting to provide unbiased estimation of treatment effects (D'Agostino, 1998).

The propensity score of an individual is formally defined as the conditional probability of being treated given his/her measured covariates. Consequentially, it is the measure of the likelihood that a person would have been treated using only their covariate scores. It can be used to stabilize the covariates in the two groups and as a result reduce the bias. Rosenbaum and Rubin (1983) showed that "it is a balancing score and can be used in observational studies to reduce bias through the adjustment methods mentioned

above. In order to estimate the propensity score, one must model the distribution of the treatment indicator variable given the observed covariates. Once estimated the propensity score can be used to reduce bias through matching, stratification (subclassification), regression adjustment, or some combination of all three."

In this work the propensity score matching method is used to reduce bias in the comparison of a case and control group.

1.2 AIM AND OBJECTIVES OF THE STUDY

The aim of the study is to reduce the impact of bias often found in observational data through the following objectives:

1. Creating covariate balance between a case and control group and hence rendering them comparable.
2. To generate a list of potential matches for each of the cases.

1.3 JUSTIFICATION FOR THE STUDY

- Comparing treatment effects on a case-control group is difficult because there are almost always baseline differences between the groups. Since observational studies are not experiments (as are randomised controlled trials) it is hard to control many external variables. Propensity-score methods are increasingly being used to reduce the impact of treatment-selection bias in the estimation of treatment effects using observational data in other countries: to date no work on propensity scores has been published in Nigeria. Moreover, though popular in cohort studies, little attention is being paid to the use of propensity scores in case-control and case-cohort studies (Mansson *et al*, 2007).

1.4 SIGNIFICANCE OF THE STUDY

- Possibility of identical comparison under some nonrandomized conditions. From this study, future researchers will gain theoretical and practical understanding of propensity score matching and its relevance to observational studies. They will be motivated to develop the skills to apply propensity scores to adjust for selection bias.
- Provision of a way to summarize covariate information about treatment selection into a scalar value.
- Its usefulness in adjusting for differences via study design. Our study has significant implications for researchers especially in the medical field. It would work for them during design consideration of their studies.

1.5 DEFINITION OF TERMS

- **Bias:** Any systematic error in the design or conduct of a study, that results in a mistaken estimate of a treatment's effect on outcome.
- **Case-Control Study:** Observational study that first identifies a group of subjects with a certain condition (treatment) and a control group without the condition, and then looks back in time (e.g. chart review) to find exposure to risk factors for the condition. This type of study is well suited for rare conditions.
- **Cohort Study:** Observational study in which subjects with an exposure of interest (e.g. hypertension) and subjects without the exposure are identified and then followed forward in time to determine outcomes (e.g. stroke).
- **Confounding:** This occurs when an investigator falsely concludes that a particular exposure is causally related to a condition without adjusting for other

factors that are known risk factors for the condition and are associated with the exposure.

- **Confounding variable:** A variable that is associated with both the exposure and outcome of interest that is not the variable being studied.
- **Control group:** A group of items/units without the condition of interest, or unexposed to or not treated with the agent of interest.
- **Covariates:** Often used simply as an alternative name for explanatory variables, but perhaps more specifically to refer to variables that are not of primary interest in an investigation, but are measured because it is believed that they are likely to affect the response variable and consequently need to be included in analyses and model building. A variable, such as age or gender, that is measured prior to the start of treatment, and hence is unaffected by the treatment, will be called a covariate.
- **Cross-Sectional Study:** Observational study that is done to examine presence or absence of a condition or presence or absence of an exposure at a particular time. Since exposure and outcome are ascertained at the same time, it is often unclear if the exposure preceded the outcome.
- **Observational study:** A study in which no intervention is made (in contrast with an experimental study). Such studies provide estimates and examine associations of events in their natural settings without recourse to experimental intervention.
- **Randomized Controlled Trial:** A special type of clinical trial in which assignment to an exposure is determined purely by chance.
- **Selection Bias:** Bias introduced by the way in which participants are chosen for a study. For example, in a case-control study using different criteria to select

cases (e.g. sick, hospitalized population) versus controls (young, healthy outpatients) other than the presence of condition can lead the investigator to a false conclusion about an exposure.

- **Total T4:** A measure of the total amount of circulating thyroxine in the blood. A high value can indicate hyperthyroidism; a low value can indicate hypothyroidism. Total T4 levels can be elevated due to pregnancy.
- **TPO:** Thyroid Peroxidase (TPO) antibodies are also known as Antithyroid Peroxidase Antibodies. These antibodies work against thyroid peroxidase, an enzyme that plays a part in the T4-to-T3 conversion and synthesis process. TPO antibodies can be evidence of tissue destruction, such as Hashimoto's disease, less commonly, in other forms of thyroiditis such as post-partum thyroiditis.
- **UIC:** The concentration of iodine in the urine. This is included in surveillance systems for iodine deficiency; it has been accepted to be the most sensitive indicator of iodine nutrition status.

CHAPTER 2

LITERATURE REVIEW

2.1 OBSERVATIONAL STUDIES

According to Mann (2003), "cohort, cross sectional and case-control studies are collectively referred to as observational studies." They are studies in which exposure is not assigned by the researcher. Mann (2003) continues thus: "Often these studies are the only practicable methods of studying various problems, for example, studies of aetiology, instances where a randomised controlled trial might be unethical, or if the condition to be studied is rare."

Cohort studies measure events in sequential order and so can be used to distinguish between cause and effect. They are used to study incidence, causes, and prognosis. Cross sectional studies are used to ascertain prevalence. They are relatively quick and easy but do not differentiate between cause and effect. Case-control studies compare groups retrospectively. They seek to pinpoint likely predictors of outcome and are useful for studying uncommon diseases or outcomes. They are frequently used to create hypotheses that can then be studied by means of prospective cohort or other studies.

An observational study draws inferences about the effect of a treatment (an exposure) on subjects where the assignment of subjects into a treatment (exposure) group versus a control group is outside the control of the investigator. The main problem with observational studies is that the allocation of subjects to treatment groups is not random. This non-random allocation creates bias and can mask cause and effect relationships or, alternatively, suggest relationships where there are none. Several methods have been developed to minimize the bias and confounding inherent with

observational studies. These methods include multivariate models, matching, stratification and propensity scoring. There are advantages and disadvantages to each method and the methods are often used in combination.

2.2 CASE-CONTROL STUDIES

In contrast with cohort and cross sectional studies, case-control studies are usually retrospective. People with the exposure of interest are matched with a control group. Retrospectively the researcher determines which individuals were exposed to the agent or treatment or the prevalence of a variable in each of the study groups. Where the outcome is rare, case-control studies may be the only feasible approach. As some of the subjects have been deliberately chosen because they have the condition in question case-control studies are much more cost efficient than cohort and cross sectional studies—that is, a higher percentage of cases per study.

Case-control studies establish the relative importance of a predictor variable relative to the presence or absence of a condition.

2.2.1 Bias in Observational Case-Control Studies

Bias may take two major forms.

- **Sampling bias** whereby the subjects with the condition may be a biased sample (for example, patients referred to a teaching hospital) or the controls may be biased (for example, volunteers, different ages, sex or socioeconomic group).
- **Observation and recall bias:** As the study assesses predictor variables retrospectively there is great potential for a biased assessment of their presence and significance by the subject or the investigator, or both.

2.2.2 Traditional Methods for Bias Reduction

They include but are not limited to:

- Matching: Matching constrains subjects in different exposure groups to have the same value for specific characteristics, these characteristics can be natural (i.e. gender) or calculated (i.e. propensity score).
- Stratification: Stratification divides the study population into a number of strata (subsets), so that subjects within a stratum share a specific characteristic and each stratum is analyzed separately.
- Multivariate models: Multivariate analyses are methods that simultaneously adjust (control) for several characteristics to estimate the independent effect of each one, these characteristics can be natural (i.e. gender) or calculated (i.e. propensity score).

These methods have flaws of their own as briefly outlined below:

- Stratification requires a stratifying variable or group of variables. Since there are likely to be several confounders between treatment (exposure) and outcomes, several stratifying variables may be required to properly balance between the treatment and comparison groups. As the number of stratifying variables increases, the number of groups becomes unmanageable, for example, a likely scenario of stratifying on five dichotomous variables (i.e. age, gender, previous immunosuppressant use, corticosteroid dose, and previous surgery) results in 32 groups. As the number of groups increases, the required sample size increases dramatically and the likelihood of adequate numbers within each strata decreases.

- Matching is a process that ‘matches’ treated subjects with comparison subjects with similar covariates. Matching does not assume a specific functional form of the treatment-outcome relationship nor does it require a homogeneous treatment effect across the population, which is advantageous since these assumptions are often unjustified (Zhao, 2005). It is often the case that several confounders exist. Therefore, matching may be required on several covariates and this may be difficult to accomplish. To address this problem matching on a multidimensional vector (such as a propensity score) may be used.

- Multivariable model risk adjustment approaches can only adjust for variables included in the model and can be effective in reducing bias when considering treatment options with similar clinical indications and risks (Concato *et al*, 2000). Observational studies comparing more diverse treatment strategies are more prone to bias since subjects who are candidates for a treatment or condition often differ in important and immeasurable ways from subjects who are not. In addition to poor performance in estimating treatment effects with relatively different treatments or conditions, multivariate models perform poorly when events are rare since several outcomes per variable are required for the reliable estimation of parameters using maximum likelihood methods (Harrell *et al*, 1996).

  Recent studies show that while results are often similar, propensity score models give treatment effect estimates that are closer to the true marginal treatment effect than multivariate regression models in which all confounders are modelled (Martens, *et al*, 2006).

2.3 PROPENSITY SCORING

Propensity scoring uses the conditional probability of being treated given the baseline covariates, the propensity score can be used to balance the covariates in different treatment groups.

2.3.1 Propensity Score Definition

With complete data, Rosenbaum and Rubin (1983) introduced the propensity score for subject i $(i=1,2,...,N)$ as the conditional probability of assignment to a particular treatment $(Z_i=1)$ versus control $(Z_i=0)$ given a vector of observed covariates, x_i:

$$e(x_i)=pr(Z_i=1 \mid X_i) \quad \text{..... (2.3.1)}$$

The propensity score is the 'crudest function' of the covariates that is a balancing score, where a balancing score, $b(X)$, is defined as 'a function of the observed covariates X such that the conditional distribution of X given $b(X)$ is the same for treated $(Z=1)$ and control $(Z=0)$ units'.

Propensity score models attempt to control confounding by adjusting, matching or stratifying the treatment effect by the propensity score. The propensity score is the probability that a subject will receive a treatment based on their pre-treatment covariates. Estimates of treatment effects with propensity score models are often similar to estimates from multivariate models (Shah *et al*, 2005 and Sturmer *et al*, 2006).

The three most common propensity score methods to balance the groups based on the measured covariates are: (1) matching, (2) stratification, and (3) regression adjustment. Matching algorithms use the propensity score to match one or more controls to each case. Stratification (or sub-classification) methods group individuals by strata based on the propensity scores. Often the boundaries of the strata are based on dividing the distribution of the entire sample into quintiles of the propensity score. Analyses are then carried out by strata. Regression adjustment techniques include the propensity score in the model as a covariate or use the propensity score as a weight. D'Agostino (1998) provides detailed reviews of these three techniques along with examples.

Matching is an attractive choice to control for bias relative to simple adjustments of the covariates made in the regression approach because regression models generally assume there is a linear relationship between the covariates and the outcome of interest. Matching does not make this assumption. Further, unlike the regression approach, the process of matching removes individuals in the control group who are a poor match to the cases. Propensity score matching has also been found to reduce bias due to case-control differences better than stratification methods based on the propensity score (Austin, 2008).

Here we illustrate the benefits of the propensity score matching approach. We used propensity score methods in a study that examines the intellectual performances of children born prematurely (cases) relative to those born at term (the controls). This is an observational study as individuals obviously could not be randomly assigned to condition (term/post-term versus pre-term). The cases and controls differ on a number of key demographic characteristics resulting in the reduced ability, because of potential

bias, to make group comparisons to look at the unique effect of prematurity on intellectual capability. Propensity score matching was used to control for potential confounding. For this study, the propensity score is defined as the probability of a baby to be pre-term, conditional on the set of measured covariates which include demographics, medical history and so on.

Debate continues on how the relationship of propensity model variables to both the treatment status and the outcome may introduce bias. In addition to the debate on which variables to include there is debate in the literature if it is better to include too many rather than too few variables. First, it is important to note that the propensity score is not used to estimate parameter coefficients or predict outcomes. The primary goal of the propensity score is to produce balance in order to reduce bias with the hopes of improving the quality of treatment effect estimates in the presence of unmeasured confounders (D'Agostino and Rubin, 2000).

Similarly, Austin *et al* (2007) noted that "the reduction in bias for propensity score models that include all measured variables was better than models that included variables associated with treatment selection alone." Austin *et al* (2007) and Heckman *et al* (1997) warn that "omitting an important variable from the propensity score will result in a biased estimation of the treatment effect." Finally, Bryson *et al* (2002) noted that, "the inclusion of non-significant variables in the propensity score will not bias the propensity-score estimates or make them inconsistent; however, it may increase their variance."

2.4 THE USE OF THE PROPENSITY SCORE TECHNIQUE

Several investigators have used the propensity score as a balancing technique and hence in reducing bias; thus giving it an important role in observational studies.

2.4.1 Matching with Propensity Scores

Often investigators are confronted with studies where there are a limited number of treated patients and a larger (usually much larger) number of control patients. A typical example (D'Agostino, 1998) is:

- A March of Dimes funded study examining the effects of post-term birth on neuropsychiatric, social and academic achievements among school-aged children (that is, 5-10 year old children). The investigators had, at the onset of the study, a collection of almost 10,000 birth records (749 treated babies (post-term) and 9421 potential controls (term) babies) with prenatal and birth history information.

Matching is a common technique used to select control subjects who are matched with the treated subjects on background covariates that the investigator believes need to be controlled.

Although the idea of finding matches seems straightforward, it is often difficult to find subjects who are similar (that is, can be matched) on all important covariates, even when there are only a few background covariates of interest. Propensity score matching solves this problem by allowing an investigator to control for many background covariates simultaneously by matching on a single scalar variable.

Rosenbaum and Rubin (1985) outline three techniques for constructing a matched sample which use the propensity score: (i) nearest available matching on the estimated propensity score; (ii) Mahalanobis metric matching including the propensity score; and (iii) nearest available Mahalanobis metric matching within calipers defined by the propensity score.

It was unfeasible financially for the investigators of the March of Dimes study to collect outcome measurement on all potential subjects that some form of sampling needed to be performed. They used the Mahalanobi's metric matching within calipers based on the propensity score method to control for more than ten background covariates. They succeeded in removing the bias between the term and post-term groups (over 74 per cent for each of the covariates that had initial standardized bias greater than 20 per cent). Another example on matching follows.

- In a study of the effects of prenatal exposures to barbiturates, Rosenbaum and Rubin (1985) used the propensity score to match 221 exposed children to 221 unexposed controls drawn from a reservoir of 7,027 potential controls. The study involved extensive follow-up of these 442 children, so it was not practical to follow all 7, 027 potential controls. The matching removed not only most of the initial bias in the 20 covariates but also most of the bias in the squares of the 20 covariates and the 190 interactions formed by multiplying two covariates.

2.4.2 Stratification (Subclassification) with Propensity Scores

Stratification (sometimes referred to as subclassification) is also commonly used in observational studies to control for systematic differences between the control and

treated groups. This technique consists of grouping subjects into strata determined by observed background characteristics. Once the strata are defined, treated and control subjects who are in the same stratum are compared directly. Many of the same problems occur in stratification as with matching when the number of covariates increases. Cochran (1965) noted that as the number of covariates increases, the number of strata grows exponentially. For instance, if all covariates were dichotomous categorical variables, then there would be 2k subclasses for k covariates. If k is large, then some strata might contain subjects from only the treated group, which would make it impossible to estimate a treatment effect in that stratum. Here again the propensity score is very useful. Because the propensity score is a scalar summary of all the observed background covariates, stratification on it alone can balance the distributions of the covariates in the treated and control groups without the exponential increase in number of strata.

Rosenbaum and Rubin (1983) presented theoretical results showing that perfect stratification based on the propensity score will produce strata where the average treatment effect within strata is an unbiased estimate of the true treatment effect. Again they assume that the treatment assignment is strongly ignorable. Rosenbaum and Rubin (1984) stated that Cochran's (1965) result, which indicated that creating five strata removes 90 per cent of the bias due to the stratifying variable or covariate, holds for stratification based on the propensity score. They state that, in fact, stratification on the propensity score balances all k covariates that are used to estimate the propensity score, and often five strata based on the propensity score will remove over 90 per cent of the bias in each of these covariates.

- Rosenbaum and Rubin (1984) studied the properties of the propensity score when used to stratify subjects in different treatment groups. In their example, the propensity score was the probability of receiving either coronary artery bypass surgery or medical therapy given by 74 different covariates. The investigators used a multi-stage procedure to find the best model for the propensity score. They found that using five strata based on the propensity score was able to substantially reduce the bias in all 74 covariates simultaneously.
- Fiebach *et al* (1990) used propensity scores to stratify patients who had received one of two possible treatments when they came to a hospital with uncomplicated chest pain. The two treatments were either admittance to a step-down unit or admittance to a coronary care unit. Covariates used to estimate the propensity scores included variables for the actual triage and independent clinical predictors for an adverse event. These clinical predictors consisted of more than 50 clinical characteristics. A stepwise procedure was used to estimate the propensity where covariates were entered into the model if they were significant at 0.50 level in a stepwise discriminant analysis.
- Stone (1995) reported an investigation whereby outcomes on 747 patients with community acquired pneumonia (CAP) who were either hospitalized (265) or ambulatory (482) were compared.

Since patients were not randomized to be either hospitalized or ambulatory, propensity scores were estimated using classification tree techniques. Patients were then assigned to one of seven strata based on their propensity score. The investigators found that there were imbalances between the two groups on 29 out of 44 baseline variables, and that

after stratification on the propensity score only 13 out of these remained significant at p = 0.05.

- In an active management of labour trial (ACT) (Berk and Newton, 1985), the investigators, in addition to the original study on the effects of active management of labour on the probability of having a Caesarean section, were also interested in determining whether the use of epidural anaesthesia was associated with Caesarean section in nulliparous women. Out of 1778 women, 1003 (56.4%) had received an epidural (which is not a random assignment).

The investigators identified 15 variables which they felt may be imbalanced between the women who received an epidural and those who did not. Propensity scores were estimated using logistic regression and the women were separated into quintiles defined by the propensity scores. The groups were more similar within each propensity score quintile than they were before stratification.

2.4.3 Regression (Covariance) Adjustments with Propensity Scores

Propensity scores can also be used in regression (covariance) adjustment. In regression adjustment, the treatment effect, τ, is estimated as

$$\hat{\tau} = (\bar{Y}_t - \bar{Y}_c) - \beta(\bar{X}_t - \bar{X}_c) \quad \text{..... (2.4.3)}$$

where the t and c indicate treatment and control groups. The effect of the background covariates is adjusted for by subtracting out the second term on the right hand side of the above equation, where b is an estimate of the regression of the responses for the treated and control groups on the background covariates. For the reasons stated above, the propensity score is a useful variable in regression adjustments, since one only has to

find the regression of the responses on the propensity scores in the treated and control groups and use this to adjust the final estimate of the treatment effect. Roseman (1994) found that "if the response surfaces in the treatment and control groups are parallel and either linear or non-linear, then the regression adjustment using the propensity scores reduces the bias in the estimate of the treatment effect."

In addition, if one stratifies and then uses regression adjustment within the strata, then this estimator of estimated treatment effect appears to be a more efficient estimator than one based on matching alone. Another approach to regression adjustment is to use a large set of background covariates to estimate the propensity score and then take a subset of these covariates and the propensity score and use them in the regression adjustment (D'Agostino, 1998).

This is the analysis performed above by the ACT investigators. The method is also analogous to performing Mahalanobis metric matching within calipers on a subset of the important covariates including the propensity score.

- In Berk and Newton (1985), an investigation to determine whether new spousal violence was influenced by whether men were either arrested or not arrested for wife-battery was reported. Here the propensity score was the probability of being arrested and 14 covariates were used to estimate it. Some of these covariates included the suspect's age, whether the victim was injured, whether the suspect was drinking, and whether the victim had called the police. The investigators then 'regressed whether or not there was new spousal violence during the follow-up (by the same offender against the same victim) on the propensity scores, separately for the arrested and not arrested group'. They found that the arrested and not arrested

groups had similar intercepts but different slopes. For the arrested group, the slope was near zero, but for the not arrested group the slope was very steep. This seemed to suggest that subjects who had a high propensity to be arrested, but were not arrested, were more likely to commit violence during the follow-up period.

- Muller *et al*. (1986) reported a study on the effects of digoxin in mortality rates in patients after myocardial infarction. Here the non-randomized treatment was digoxin. The investigators calculated an 'imbalance risk score', which appears to be a propensity score, based on 19 covariates. The covariates in this model included the subject's heart rate, age, and whether the subject had beta-blockers in the previous three weeks. Cox proportional-hazards regression was used to determine the association between digoxin therapy and survival 'taking into account the effects of baseline prognostic factors'. It appears that they 'took into account' the baseline differences by including a propensity score as a covariate in their model in order to adjust their final treatment effect.

2.5 LOGISTIC REGRESSION

2.5.1 Introduction

When we wish to model a binary dependent variable, then an appropriate analysis is often logistic regression. Logistic regression is a generalisation of the chi squared test to examine the association of a binary-dependent variable with one or more independent variables which can be binary, categorical (more than two categories) or continuous. Logistic regression is also useful for analysing case-control studies. Matched case-control studies require a particular analysis known as conditional logistic regression (Campbell, 2001).

2.5.2 The model

Logistic regression model generalizes contingency table methods for binary outcomes. The dependent variable can be described as an event which is either present or absent (sometimes termed 'success' and 'failure'). Thus an event might be the presence of a disease in a survey or cure from disease in a clinical trial. We wish to examine factors associated with the event. Since we can rarely predict exactly whether an event will happen or not, what we in fact look for are factors associated with the probability of an event happening (Campbell, 2001).

There are two scenarios to consider; the independent variables may either be all categorical whereby one can calculate the proportion of subjects for whom an event happens, or at least one of them is continuous.

The population parameters are modelled. If we consider the categorical grouped case first, denote the population probability of an event for a cell i by p_i. This is also called the 'expected' value.

Thus, for an unbiased coin, the population or expected probability for a 'head' is 0.5. The dependent variable y_i is the observed proportion of events in the cell (say the proportion of heads in a set of tosses) and we write $E(y_i) = p_i$ where E denotes 'expected value'. Also if an event has probability i, then the odds for that event are $\frac{p_i}{(1-p_i)}$ to 1. Thus the odds of a head to a tail are 1 to 1. The model is

$$\log_e\left(\frac{p_i}{1-p_i}\right)=\log it(p_i)=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+...+\beta_p X_{ip} \quad \text{..... (2.5.2)}$$

where the independent variables are $X_{i1},...,X_{ip}$.

The term on the left-hand side of the equation is the log odds of success, and is often called the logistic or logit transform. Suppose we have only one covariate x, which is binary and simply takes the values 0 or 1. Then the odds ratio associated with x and y is given by exp (β). If x is continuous then exp (β) is the odds ratio associated with a unit increase in x.

The main justification for the logit transformation is that the odds ratio is a natural parameter to use for binary outcomes, and the logit transformation relates this to the independent variables in a convenient manner. It can also be justified as follows. The right-hand side of equation is potentially unbounded, that is can range from minus to plus. On the left-hand side, a probability must lie between 0 and 1. An odds ratio must lie between 0 and ∞. A log odds ratio, or logit, is unbounded and has the same potential range as the right-hand side of the equation.

Note at this stage that the observed values of the dependent variable are not in the equation. They are linked to the model by the Binomial distribution. Thus in cell i if we observe y_i successes in n_i subjects, we assume that the y_i are distributed Binomially with probability p_i. The parameters in the model are estimated by maximum likelihood.

2.5.3 Uses of Logistic Regression

1. As a substitute for multiple regression when the outcome variable is binary in cross-sectional and cohort studies and in clinical trials. Thus we would use

logistic regression to investigate the relationship between a causal variable and a binary output variable, allowing for confounding variables which can be categorical or continuous.

2. As a discriminant analysis, to try and find factors that discriminate two groups. Here the outcome would be a binary variable indicating membership to a group. For example one might want to discriminate men and women on psychological test results.
3. To develop prognostic indicators, such as the risk of complications from surgery.
4. To analyse case-control studies and matched case-control studies.

2.6.1 PRETERM BIRTH

Preterm birth in humans refers to the birth of a baby of less than 37 weeks gestational age. **Premature birth**, which is commonly used as a synonym for preterm birth, refers to the birth of a baby before the developing organs are mature enough to allow normal postnatal survival.

The cause for preterm birth is in many situations elusive and unknown; many factors appear to be associated with the development of preterm birth, making the reduction of preterm birth a challenging proposition. Maternal background is one of the factors that have been identified as linked to a higher risk of a preterm birth. This can be in the form of age at the upper and lower end of the reproductive years, be it more than 35 or less than 18 years of age; height and weight; low BMI; type of work and physical activity have given conflicting results, but it is opined that stressful conditions, hard labour, and long hours are probably linked to preterm birth.

Obesity does not directly lead to preterm birth; however, it is associated with diabetes and hypertension which are risk factors by themselves.

Marital status is associated with risk for preterm birth because pregnancy outside of marriage is more associated with adverse outcomes.

There are factors during pregnancy such as multiple pregnancies (twins, triplets, etc.); high blood pressure, pre-eclampsia, maternal diabetes, asthma, thyroid disease, and heart disease are some maternal medical conditions that increase the risk of preterm birth, and often labour has to be induced for medical reasons. Babies with birth defects are at higher risk of being born preterm.

Presence of anti-thyroid antibodies is associated with an increased risk preterm birth with an odds ratio of 1.9 and 95% confidence interval of 1.1–3.5.

The Nigerian cultural method of abdominal massage has been shown to result in 19% preterm birth among women in Nigeria, plus many other adverse outcomes for the mother and baby. This ought not to be confused with massage conducted by a fully trained and licensed massage therapist or by others trained to provide massage during pregnancy, which has been shown to have numerous positive results during pregnancy, including the reduction of preterm birth, less depression, lower cortisol, and reduced anxiety.

2.6.1 Consequences of Preterm Births

Premature babies face an increased risk of lasting disabilities, such as mental retardation, learning and behavioural problems, cerebral palsy, lung problems and vision and hearing loss. "Two recent studies suggest that premature babies may be at increased risk of symptoms associated with autism (social, behavioural and speech problems). (Limperopoulos *et al*, 2008; Schendel and Bhasin, 2008)

Studies also suggest that babies born very prematurely may be at increased risk of certain adult health problems, such as diabetes, high blood pressure and heart disease (Hovi *et al* 2007).

Many children will adjust well during childhood and adolescence, although a large study that followed children born between 22 and 25 weeks found some alarming results. When the children were 6 years old it was found that 46 percent had severe or moderate disabilities such as cerebral palsy, vision or hearing loss and learning problems; 34 percent were mildly disabled while 20 percent had no disabilities. 12 percent had disabling cerebral palsy. Some of the complications related to prematurity may not be apparent until years after the birth. A long-term study demonstrated that the risks of medical and social disabilities extend into adulthood and are higher with decreasing gestational age at birth and include cerebral palsy, mental retardation, disorders of psychological development, behaviour, and emotion, disabilities of vision and hearing, and epilepsy. Standard intelligence tests showed that 41 percent of children born between 22 and 25 weeks had moderate or severe learning disabilities when compared to the test scores of a group of similar classmates who were born at full-term. It is also shown that higher levels of education were less likely to be obtained

with decreasing gestational age at birth. People born prematurely may be more susceptible to developing depression as teenagers.

CHAPTER 3

METHODOLOGY

3.0 INTRODUCTION

Observational research is often used in studies such as that of studying the relationship of status at birth and future intellectual capabilities. It allows scientists to examine relevant factors that may provide the framework for future intervention research. Studying the effects of prematurity at birth on a child's intellectual capacity requires an observational rather than experimental design. However, in observational studies such comparisons are challenging because infants born at full-term or post-term and those born preterm may differ in their distributions on key demographic and other comorbid conditions related to their status at birth.

In experimental studies, random assignment to condition ensures that on the average there should be no systematic differences between groups on relevant factors (D'Agostino, 1998).Thus, the randomization process balances group differences on both measured and unobserved variables. However, in observational studies, there is no random assignment and group differences on confounding variables may potentially have significant effects on the results. It is important that the term/post-term and preterm babies are as similar as possible on the characteristics other than their status at the time of study in order to reduce the possibility of confounding when examining future differences in intellectual performance. Theoretically, if T and C represent treatment and control groups, respectively, we can let

$Y_i^T \equiv$ intellectual index (outcome) of a baby born preterm, and

$Y_i^C \equiv$ intellectual index (outcome) of a baby born otherwise.

Then Y_i^T - Y_i^C ≡ effect of being born preterm for a given baby. The problem here is that a baby can only belong to one group.

Define E [Y_i^T - Y_i^C] ≡ mean effect of being born preterm. Looking at the average effect of preterm birth on the babies, the mean difference D between the babies in the treatment group and those in the control group is given as

D = E [Y_i^T |T] – E [Y_i^T|C] = E [Y_i^T |T - Y_i^C |C]

= {E [Y_i^T − $E[\ Y_i^C$|T]} + {E [Y_i^C|T] – E [Y_i^C |C]}

The term τ_i = {E [Y_i^T − $E[\ Y_i^C$|T]} ≡ average difference the treatment makes among treated babies;

∂_i = {E [Y_i^C|T] – E [Y_i^C|C]} ≡ selection bias due to systematic differences between the T babies and C babies.

Define S ≡ preterm birth (T = 1, C = 0) and X = vector of measured covariates.

Let p*(X)* ≡ Prob[S =1| *X]*. Then the term

{E [Y_i^C|T] – E [Y_i^C|C]} implies that {E [Y_i^C| p*(x), S*] – E [Y_i^C| p*(x)*, C]} = 0.

In order to investigate the effect of S, the strategy is to find ways of reducing the bias term. By using the estimate of the propensity score and then comparing observations on babies that have similar propensity score or logit of propensity score, then the observable selection bias is reduced and the treatment effect τ is isolated from *D*.

Researchers have relied on statistical approaches to help reduce the bias due to group differences in observational studies. However, these approaches have limitations. For example, traditional methods to control for group differences such as matched sampling, stratification, and covariate adjustment are often limited by the number of variables that can be adjusted because of sample size concerns (D'Agostino, 1998). The

main focus of this work is to illustrate the strength of this bias reduction method, propensity score matching (given its potential for improved precision and a reduction in bias with propensity score models over multivariate regression), within a case-control study of preterm and term/postterm babies. This approach is highlighted for its utility in selecting term/postterm babies (controls) to compare to those preterm babies (cases), matching on a number of relevant factors of interest.

Propensity score matching simply uses the traditional framework of matching two groups to make them comparable, but matches them on a single indicator, the propensity score, rather than multiple variables. When matching, controls from the term/postterm group are selected who have similar propensity scores to those in the preterm group (cases). The goal is a dataset of cases and controls with similar characteristics on all key variables that were used to define the propensity scores. The propensity score is defined as the probability of being in the case group, given the individual's level on the covariates included in the model (Rosenbaum and Rubin, 1983).

The propensity score is often estimated using a logistic regression model because logistic regression makes no assumptions about the distributions of the covariates on the dichotomous outcome (D'Agostino, 1998).

A single propensity score is estimated for every individual in the study, both cases and controls. This propensity score is then used to adjust for the differences between the two groups on the observed covariates in the study. Thus, the propensity score is often thought of as a balancing score allowing researchers to control for a large number of

background covariates simultaneously based on a single number (Rosenbaum and Rubin, 1983).

The final study design shall be a propensity score matched prospective cohort study built out of a retrospective case-control study. Preterm babies shall be matched with term/postterm babies for follow-up on their intellectual achievements based on their propensity scores.

3.1 PRELIMINARY STATISTICAL ANALYSIS

Before the logistic regression analysis, there is a need to measure the association between the response variable which is the status of a baby and the explanatory variables. The measure of association used between the response variable and the ordinal categorical variables was the Kendall's tau-c, the Phi for the nominal categorical variables and the Eta for the continuous variables. Testing of association was done to determine variables that are related to baby status.

3.1.1 Correlation Coefficients

A correlation coefficient describes how closely two variables are related, that is the amount of variability in one measurement that is explained by another measurement. The range of a correlation coefficient is from −1 to +1, where +1 and −1 indicate that one variable has a perfect linear association with the other variable and that both variables are measuring the same entity without error. In practice, this rarely occurs because even if two instruments are intended to measure the same entity both usually have some degree of measurement error.

A correlation coefficient of zero indicates a random relationship and the absence of a linear association. A positive coefficient value indicates that both variables increase in value together and a negative coefficient value indicates that one variable decreases in value as the other variable increases. It is important to note that a significant association between two variables does not imply that they have a causal relationship. Also, a correlation coefficient that is not significant does not imply that there is no relationship between the variables because there may be a non-linear relationship such as a curvilinear or cyclical relationship.

Correlation coefficients are rarely used as important statistics in their own right. An inherent limitation is that correlation coefficients reduce complex relationships to a single number that does not adequately explain the relationship between the two variables. Another inherent problem is that the statistical significance of the test is often over-interpreted. The *P* value is an estimate of whether the correlation coefficient is significantly different from zero so that a small correlation of no clinical importance can become statistically significant, especially when the sample size is large. In addition, the range of the data as well as the relationship between the two variables influences the correlation coefficient.

Kendall's τ (tau) is used for correlations between two categorical or non-normally distributed variables. In this test, Kendall's τ is calculated as the number of concordant pairs (P) minus the number of dis-concordant pairs (Q) divided by the total number of pairs. Kendall's tau-c is suitable when the two variables have an unequal number of

categories. It is calculated as $\tau_C = \frac{2m(P-Q)}{n^2(m-1)}$, where m = min (row, column) and n = sample size.

Phi provides a test of statistical significance and also provides information about the strength of the association between two categorical variables. It is appropriate for a 2 x 2 cross-tabulation. Values of larger than or equal to +/- 0.5 indicate strong relationships. Phi is given as $\phi = \sqrt{\frac{\chi^2}{N}}$, where χ^2 is the usual chi-square statistic for testing the independence of the two variables and N is the sample size.

Eta is used when one variable is nominal and the other is approximately normal or continuous. It is a measure of association that ranges from 0 to 1, with 0 indicating no association between the row and column variables and values close to 1 indicating a high degree of association.

3.2 Propensity Scores

The propensity scores for our study are the probabilities of babies being born preterm given our chosen independent variables.

3.2.1 Propensity Score Variable Identification

Some clinically relevant covariates were selected, based on information available. In addition, a few demographic variables were included in the model to ensure general consideration of possible confounders.

3.2.2 Variable Selection for Propensity Score Model

Some authors suggest including all variables thought to be related to the outcome in the propensity score model, whether or not they are related to treatment (Rubin and Thomas, 1996). Others suggest that only variables that influence simultaneously treatment and outcomes should be included (Caliendo & Kopeinig, 2005; Perkins *et al*, 2000 and Bryson *et al*, 2002).

Since the propensity score serves only as a device to balance the observed distribution of covariates across treated and comparison groups, the propensity score covariate coefficients are not of direct importance to estimating treatment effects per se. Therefore, the usual reasons to reduce the number of variables in a model may not apply to propensity score modelling. In addition to reduced concerns about model parsimony, there are other unique situations to propensity score model fitting; for example, over-fitting is not a concern, since model generalizability is not relevant to bias reduction.

Therefore, the number of covariates, the number of functional forms and the number of interactions can be much larger than would be routinely considered in regression analyses. Rubin and Thomas (1996) specifically recommend against `trimming' models in the name of parsimony, they argue that a variable should only be excluded from analysis if there is consensus that the variable is either unrelated to the outcome or not a proper covariate. If there are doubts about these two points, they explicitly advise to include the relevant variables in the propensity score estimation.

There is considerable evidence to suggest that in general it may be better to include 'potentially useless' covariates rather than undergo the risk of excluding relevant ones (Cavuto *et al*, 2006).

In contrast to a pre-planned variable selection strategy, Rubin (2004) offers the following insight:

> "If this balance is achieved in an observational study—that is, if the treatment and control groups have very similar distributions of the observed covariates within blocks (subclasses, matched pairs etc.) of the propensity score (perhaps crossed by blocks on critical covariates)—then it really makes no difference, for estimation of effects controlling for these covariates, as to how this balance was achieved."

While including all covariates reduces the chances of imbalance, bias and missing potential confounders, there is a statistical price to pay as this strategy may reduce the proportion of subjects who are able to be matched and it may increase the variance of the propensity score (Caliendo & Kopeinig, 2008).

In our study, we elected to construct the propensity score using a model which included only those variables that were found to be related to baby status at birth.

3.2.3 Propensity Score Modelling

Our base propensity model was constructed using logistic regression with some medical and prematurity related covariates identified from questionnaires administered and results from urine and blood tests. Logistic regression was chosen since, according to Ankarali *et al*, (2009), "it does not require any assumptions relating to covariate distribution."

3.3 LOGISTIC REGRESSION AND PROPENSITY SCORES

A propensity score is obtained as the probability score of an individual from the fitted simple logistic regression model of the two groups as the dependent variable to the set of covariates. Logistic regression is applied when the dependent variable is dichotomous. In this study, the dependent variable is S = Status of baby (1 = pre-term, 0 = term/postterm). Let P(X) = Prob[S =1\ X]. The logistic regression is defined as

$$P(X)=PS=\frac{\exp(\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p)}{1+\exp(\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p} \quad \text{..... (3.3a)}$$

where: PS = Propensity Score, β_0 = constant of the equation; and ßi = coefficients of the covariates Xi;

Alternatively, the equation above can take on a linear form through logit transformation, also known as logit regression equation,

$$L(X) = \ln\left(\frac{PS}{1-PS}\right) = \beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p \quad \text{..... (3.3b)}$$

The propensity score is the probability of a baby to be in the pre-term group given his observed covariates X. This can be computed by substituting the values of the covariates in the logistic regression equation.

3.3.1 Elements of the Propensity Score

Data Collection

Even though the original dataset contained a lot of variables, the independent variables chosen for the study are only those found to be significantly correlated with the chosen dependent variable.

Dependent Variable for estimating the propensity scores:

Status: pre-term, full-term/post-term.

Independent Variables: These are the main concerns for this study.

Variables related to the child:

Socio-demographic:

Sex (Male/Female): A human being is either a male or a female.

Birth weight (Low/Normal or Macrosomia): The weight of a baby at birth. A weight of less than 2.5 kilogram is considered low.

Gestational age at delivery/birth: The age in weeks of a baby at the time of its delivery. Babies born before the thirty-seventh week are termed premature, thereafter up to the fortieth week they are full-term babies. Post-term babies are those born beyond 42 weeks.

Variables related to the mother:

Medical History/Symptoms: **TPO** (-ve/+ve), urinary iodine concentration **(UIC)** in µg/l and total T4 **(TT4)** nmol/l. These have been explained earlier on under the list of terminologies.

Socio-demographic:

Body mass index (Underweight/Normal or Overweight): Body mass index (BMI) is a mathematical ratio of height to weight that can be linked with body composition (or body fat percentage) and with indices of health risk. A BMI of 18.5 to 24.9 is considered normal and people with anything lower than that are seen as underweight. People with a BMI of 25 to 29.9 are considered overweight, and people with a BMI of 30 or above are considered obese. For this study, BMI was converted from a continuous to a categorical variable.

Parity (Primiparous/Multiparous): Parity refers to the condition of having given birth or the number of children a woman has given birth to. A primiparous mother has given birth only once while a multiparous one has delivered a multiple number of times. Parity was converted from a discrete variable to a categorical one.

To assess balance, the independent samples t-tests (and the χ^2 tests as appropriate) were used to test the effect of each covariate on the treatment status. Imbalance between the two groups was denoted by p-values of less than 0.05 and percentage standardized bias of greater than 10% in absolute value. The primary goal of the propensity score modelling is to balance covariates between the two groups with respect to the propensity score, however, model fit statistics were evaluated to gain insight into the model. The first statistic was the Pseudo R-Square statistic (Cox and Snell and Nagelkerke statistics; they are approximations to the r-square statistic because such cannot be exactly computed for logistic regression models) to see what proportion of variation is explained by the model. Another statistic is the Wald statistic which is a chi-square based test that assesses the contribution of each parameter in the model. The Hosmer-Lemeshow test (also a chi-square-based test) which assesses whether the model fits the data adequately was also used.

3.4 STUDY POPULATION

The data for this study is secondary obtained from Ahmadu Bello University Teaching Hospital in Zaria. From an original dataset of 414 babies used to study nutritional iodine status, 13 preterm babies were chosen as cases and 203 full-term/post-term babies as the control group. They are all potential subjects from the original population.

For our study, we assumed that the researcher (original owner of the data) might be interested in whether a baby's status at birth was associated with his/her future intellectual capabilities. This intellectual capability is however out of the scope of this study; the main concern of the study is finding the best possible sample from which to measure outcome (intellectual capability) in the future. In any case, none of the babies had attained the age when the outcome could be measured at the time of this study.

Exclusion criteria

Any subject with at least one single missing value on any of the covariates was excluded, and so were babies that died at birth.

3.5 ANALYSIS SELECTION

There are several different approaches for using the propensity score to estimate treatment effects, including:

i. Covariate adjustment using the propensity score.
ii. Stratification on the propensity score.
iii. Propensity-score matching.

Regression or covariate adjustment using treatment status and the estimated propensity score assumes that the correct propensity score model has been specified and requires sufficient overlap in the propensity score between subjects in the two groups being compared. We decided against covariate adjustment due to the stringent requirement of the assumption that the propensity score is correct and we predicted that there would not be much overlap in the propensity score between the two groups.

Stratification requires adequate number of term/postterm and preterm subjects in each stratum. We expect that there would be a smaller number of treated cases than there would be of non-comparable control cases within the quintiles of the stratum, a situation which makes stratification on the propensity score difficult.

Published empirical and theoretical research has demonstrated that matching on the estimated propensity score can result in a greater reduction in treatment-selection bias than does stratification on the estimated propensity score (Austin *et al*, 2007 and Austin and Mamdani, 2006).

Given the improvement in bias reduction and the expected low proportion of pre-term babies in some propensity score quintiles we elected to match subjects based on the propensity score.

3.5.1 Matching

Matching was used to select a comparison group from term/postterm babies that is similar to the group of preterms with respect to the distribution of the propensity score (and potentially other confounders). Below are some major considerations when implementing a strategy to create a matched sample:

i. matching metric,
ii. matching algorithm,
iii. matching with or without replacement, and
iv. number of comparisons per treated subject.

Within the context of matching, bias is the difference between the baseline covariates in the treatment and comparison groups. A small bias after matching indicates that the quality of matching is good (Smith and Todd, 2005).

3.5.2 Matching Metric

The matching variable was the logit of the propensity score, which is an unbounded continuous variable with values ranging from minus to plus. Unlike dichotomous or categorical values which can be matched exactly, a metric must be defined to evaluate the 'closeness' of a continuous variable between the treatment and the comparison group. When matching on a single variable, the simple absolute distance (or nearest neighbour) is a common choice. Matching on several variables requires more complex matching metrics such as the Euclidean distance, the Mahalanobis distance, or the sum of the absolute (or squared) difference of the ranks of the matching variable (Bergstralh *et al*, 1996 and Rosenbaum, 1989).

We used the simple absolute distance (nearest neighbour) because we used a single main matching variable (the logit of the propensity score).

3.5.3 Matching Variable

Both the propensity score and its logit were used to avoid the problems of more than one control with the exact same shortest absolute distance to a case. The first variable considered was however the logit.

3.5.4 Matching Algorithm

Matching algorithms can be classified into two general categories; local optimal or global optimal. Local optimal algorithms, often called 'greedy' algorithms, are methods that make optimal decisions at each step without attempting to make the best overall decision. Global optimal algorithms minimize the overall distance between the treatment group and their matched controls. We used global optimal algorithms as it

always produces similar or better matches than local optimal algorithms (Rosenbaum, 1989).

Since the matching algorithm can only optimize matching and does not necessarily produce appropriate matches a caliper was added as a precaution against inappropriate matching. Calipers between 0.20 to 0.60 standard deviations have been shown to remove close to 90% of the bias associated with the matching variable (Gu and Rosenbaum, 1993).

From Rosenbaum and Rubin (1985a), "The combination of optimal matching and the use of calipers produce matched groups with fewer imbalances." We used a caliper of 0.3 for our own study.

3.5.5 Matching Structure

Choosing the number of comparison units to match with each treatment unit is dependent on the dataset. By using a single comparison unit for each treatment unit the smallest distance between the treatment and comparison units is ensured. By using more comparison units, one increases the precision of the treatment effect estimates, but at the cost of increased bias. Recent work has shown that variable matching (i.e. where the number of comparison units is not fixed) can produce substantially greater bias reduction with only a slight increase in variance when compared to a fixed number of comparison units (Ming and Rosenbaum, 2000). We performed single one to one matching.

3.5.6 Replacement of Comparison Subjects

Matching with replacement will minimize the global distance between the term/postterm and preterm groups since each preterm can be matched to the nearest comparison subject, even if a comparison subject is matched more than once. This reduction in global distance results in bias reduction but may increase variance and reduce the precision of the treatment effect estimates (Rosenbaum and Rubin, 1985). Several authors note that matching without replacement is sensitive to the order in which the treatment units are matched (Dehejia and Wahba, 2002).

However, these criticisms are not valid for optimal matching algorithms using continuous variables such as the propensity score since it is unlikely that there would be two subjects with the exact same propensity score. Given the small number of participants in our study and the expectation that there would be a small number of preterm babies (estimated at 6.02% of the study population) we assumed that there would be sufficient number of relevant comparison subjects to choose from to ensure good matches without replacement.

3.5.7 Nearest-Neighbour Matching on Estimated Propensity Score

This is the easiest in terms of computational considerations. It consists of randomly ordering the treated and control subjects, then selecting the first treated subject and finding the control subject with closest propensity score. Both subjects are then removed from consideration for matching and the next treated subject is selected. Rosenbaum and Rubin (1985b) suggest using the logit of the estimated propensity score to match because the distribution of the logit is often approximately normal. The procedure is as follows:

1. The cases are randomly ordered and the first of them is selected.

2. All the control subjects with the closest logit of the propensity score or within a determined caliper of the case subject's estimated logit of the propensity score are then selected. The caliper chosen for our study was 0.30 standard deviations above or below the logit of the propensity score.

3. The next step is to calculate nearest distance between these controls and the case subject. The closest control subject and the case subject are then removed from the pool, and the process is repeated until the last case finds its match in the control group.

3.6 ASSESSING THE MATCHED DATA

Once we estimated the propensity scores, the nearest neighbour matching was performed. We then calculated descriptive statistics, re-run t-tests and χ^2 tests as appropriate and obtained standardized percentage differences as before.

It is important to realize that, since these comparisons involve only covariates and not outcome variables, there is no chance of biasing results in favour of either cases or controls through the selection of matched controls.

3.6.1 The Standardized Bias

The percentage standardized difference for incomplete one–to-one matching was used to compare the balance of baseline covariates between preterm babies and term/post-term babies in the matched set. The percentage standardized difference, also known as the standardized bias, is the mean difference as a percentage of the average standard deviation; it compares the difference in the means between two groups in units of the pooled-standard deviation. It is a suitable indicator to assess the distance in marginal

distribution of the covariate as suggested by Rosenbaum and Rubin (1985a). The standardized bias is computed as:

$$d = \frac{100(\bar{x}_T - \bar{x}_C)}{\sqrt{\frac{S_T^2 + S_C^2}{2}}} \quad \text{..... (3.6.1a)}$$

where: $\bar{x}_T$ = mean of the treatment group; $\bar{x}_C$ = mean of the control group, and

S^2_T = variance of the treatment group; S^2_C = variance of the control group

For binary data, the standardized bias is computed as:

$$d = \frac{100(P_T - P_C)}{\sqrt{\frac{P_T(1-P_T) + P_C(1-P_C)}{2}}} \quad \text{..... (3.6.1b)}$$

where: P_T = proportion of the covariate in the treatment group, and P_C = proportion of the covariate in the control group.

The advantage of the standardized difference over t-tests and other statistical tests of hypothesis is that the standardized difference is not influenced by sample size. Standardized differences represent residual biases in the baseline covariates; standardized differences greater than 10% in absolute value indicate serious imbalance (Austin *et al*, 2007).

3.6.2 Bias Reduction

The reduction due to bias is computed as:

$$BR = 100(1 - B_M/B_O) \quad \text{..... (3.6.2)}$$

where: $B_M = \frac{100(\bar{x}_{MT} - \bar{x}_{MC})}{\sqrt{\frac{S_{MT}^2 + S_{MC}^2}{2}}} \equiv$ standardized bias after matching, and

$$B_O = \frac{100(\bar{x}_{OT} - \bar{x}_{OC})}{\sqrt{\frac{S_{OT}^2 + S_{OC}^2}{2}}} \equiv \text{standardized bias before matching.}$$

where: subscript M denotes after matching; O denotes before matching.

Note: Most empirical studies show that a bias reduction of 3 to 5% is seen as sufficient.

3.6.3 Test on Means and Proportions

According to Caliendo and Kopeinig (2005) "the two sample t-test can also be used to check if there are significant differences in covariate means for both groups. Before matching, differences between the groups are expected; but after matching, the covariates should be balanced in both groups and hence no significant differences should be found. The t-test might be preferred if the evaluator is concerned with the statistical significance of the results." In the same way, the chi-square test is used to examine the effect of categorical covariates.

3.7 ANALYSIS

The descriptions of the continuous variables are expressed as mean ± standard deviation and the dichotomous and qualitative variables as percentages.

3.7.1 Software

Propensity score development and statistical analyses was conducted using SPSS 16 (Copyright (c) SPSS Inc., 1989 - 2007) on a Windows Vista platform (Microsoft, Seattle, WA). Standardized differences were calculated using Microsoft Excel Version 2007 on a Windows Vista platform (Microsoft, Seattle, WA). Matching was done with the Microsoft Excel Version 2007.

CHAPTER 4

DATA ANALYSIS AND INTERPRETATION

4.1 CORRELATIONS WITH BABY STATUS

Testing of association was done to determine variables that were related to the status of a baby at birth.

4.1.1 Correlation of Status with Covariates

Correlation analysis was performed to identify the variables that were affected by Status. There were varying degrees of relationships out of which four were found to be significant as presented in Table 4.1. Gestational age at birth is perfectly correlated to status (they are basically the same thing expressed differently). Sex and TPO have weak but significant relationships with status. The tests for correlation suggest no significant associations between status and either birth weight, BMI

Table 4.1: Correlation of Status with Covariates

Covariate	**Correlation Coefficient**
Gestational age at birth	1.000**
Birth Weight	0.033
Sex	0.139*
Urinary iodine concentration	0.787*
Body Mass Index	0.043
Total T4	0.979
TPO	0.196**
Parity	0.064

**Significant at 1%, *Significant at 5%

4.2 LOGISTIC REGRESSION ON BABY'S STATUS

After the testing for association, the Binary Logistic procedure of SPSS 16 was used in building the logistic regression model using all the variables except Gestational Age.

Table 4.2a: Logistic Regression on baby's status

Covariate	B	Std. Error	Wald	P-value.	Exp(B)	95% Confidence Interval for Exp(B)	
						Lower Bound	Upper Bound
Intercept	2.240	1.623	1.906	0.167	9.393		
UIC	-0.009	0.004	4.434	0.035	0.991	0.984	0.999
Total T4	-0.313	0.122	6.571	0.010	0.731	0.576	0.929
TPO = –ve	-1.810	.864	4.389	.036	0.164	0.030	0.890
Sex = Female	0.470	.812	0.336	0.562	1.601	0.326	7.855
BMI = Underweight	0.764	.959	.634	0.426	2.146	0.0328	14.056
Birth-weight = Low	5.828	1.769	10.854	.001	339.530	10.599	10876.793
Parity = Primiparous	1.393	.745	3.494	.062	4.027	0.935	17.350

Table 4.2a shows the estimates, standard error, the p -values and the confidence interval of the propensity model for baby status at 5% level of significance.

The model goes thus:

$$PS = \frac{e^{\left(2.240 - 0.009X_1 - 0.313X_2 + 1.81X_3 + 0.47X_4 + 764X_5 + 5.828X_6 + 1.393X_7\right)}}{1 + e^{\left(2.240 - 0.009X_1 - 0.313X_2 + 1.81X_3 + 0.47X_4 + 764X_5 + 5.828X_6 + 1.393X_7\right)}}$$

where X_1 = UIC, X_2 = TT4, X_3 = TPO-ve, X_4 = female, X_5 = underweight, X_6 = low birth-weight, X_7 = primiparous and TPO-ve, female, underweight mother, low birth-weight and primiparous are each coded 1.

The Pseudo R^2 values for the model are 0.167 (Cox and Snell) and 0.458 (Nagelkerke) meaning that, at most, 45.8% of variation between the two groups is explained by the model (see Table 4.2b below).

Table 4.2b: Model Summary and Goodness-of-fit Tests

	Omnibus test of model coefficients	Hosmer and Lemeshow Test	Cox & Snell R Square	Nagelkerke R Square
Chi-square	39.533	2.312	0.167	0.458
df	7	8		
P-value	<0.0001	0.970		

The significance of the omnibus test implies that the model is better than none at all. The Hosmer-Lemeshow statistic has a p-value greater than 0.05 (p=0.970) meaning that the data is consistent with the model (i.e. the model fits the data). According to both the Wald statistics in Table 4.2a, the variables Sex, BMI and Parity do not contribute significantly to the model.

The parameters **B** in a logistic model are not as straightforward as in a linear model; they are only convenient for testing the usefulness of predictors but **Exp(B)** are easier to interpret.

The model, amongst other things, implies that each extra unit of UIC decreases the odds of having a preterm baby by a multiplicative factor of 0.991 (see **Exp (B)** column on the UIC row), and that the odds of a preterm birth by a mother with negative TPO is 0.164 less than that of a mother with a positive TPO all other things being equal. Similar deductions can be made with the other variables. Odds here can be defined as

$$\text{Odds}(\text{preterm}) = \frac{\text{P}(\text{preterm})}{\text{P}(\text{full}-/\text{post}-\text{term})} = \frac{\text{P}(\text{preterm})}{1-\text{P}(\text{preterm})} \quad \text{and}$$

$$\text{P}(\text{preterm}) = \frac{\text{Odds}(\text{preterm})}{1+\text{Odds}(\text{preterm})}.$$

According to the model, 15.4% (only 2) of the preterm babies have been classified correctly, 98.5% of the controls were correctly classified and overall 93.5% of the subjects have been correctly classified (see Table 4.2c). The table gives the practical results of employing the logistic regression equation obtained. Each case is assigned to the class where his/her propensity is higher. The Null model classifies 94% of the subjects in the control category.

Table 4.2c: Classification Table

	Predicted		
Observed	Treated	Control	Percent Correct
Treated	2	11	15.4%
Control	3	200	98.5%
Overall Percentage	2.3%	97.7%	93.5%

4.3 GROUP COMPARISONS

Table 4.3 shows the descriptive statistics of the unmatched covariates. There were significant differences between the covariates in the cases and controls. Birth weight had the least standardized bias. All the p-values indicate that there are differences on the means of the covariates. All of the covariates have large standardized percentage differences indicating imbalance between the case and control groups. The goal of matching is to reduce the differences between these two groups.

Table 4.3: Group comparisons before matching

Continuous Covariates	**Cases (N = 13)**		**Control Group (N = 203)**		**P-value (t-test)**	**Standardized Difference in %**
	Mean	**SD**	**Mean**	**SD**		
Gestational age at birth (weeks)	34.62	0.65	40.20	2.37	<0.0001	321.11
UIC (µg/l)	122.5	113.4	224.1	161.1	0.008	72.93
Total T4 (nmol/l)	9.68	3.29	12.73	3.43	0.002	90.75
Propensity score	0.33	0.23	0.04	0.11	0.001	160.86
Logit of the propensity score	-0.91	1.28	-4.68	1.96	<0.0001	227.75
Ordinal/Nominal Covariates	**N = 13**	**%**	**N= 203**	**%**	**P-value (χ^2- test)**	**Standardized Difference in %**
TPO (-ve)	9	69.2	188	92.6	0.019	62.33
Sex (Female)	10	76.9	97	47.8	0.08	63.06
Parity (Primiparous)	**7**	53.8	52	25.6	0.048	60.25
Birth Weight (Low)	2	15.4	2	1	0.019	54.43
Body Mass Index (Underweight)	3	23.1	8	3.9	0.021	58.31

Propensity scores were computed by substituting the corresponding values of the covariates on the logistic regression equation. Logits of these scores were computed and randomized. A list of potential matches was obtained for the cases (see Appendix

I). The nearest-neighbour matching was then performed and Table 4.4 highlights the twelve matches obtained. Matches with respect to the propensity scores coincided with those made using the logits.

Table 4.4: Propensity Score Matching

Case	Propensity score	Logit of the PS	Control	Propensity score	Logit of the PS
390	0.05	-3.02	155	0.05	-2.99
326	0.06	-2.69	381	0.06	-2.71
223	0.12	-2.05	260	0.11	-2.08
411	0.16	-1.66	333	0.16	-1.70
405	0.18	-1.51	363	0.18	-1.51
42	0.23	-1.20	203	0.24	-1.13
415	0.31	-0.81	289	0.32	-0.76
286	0.37	-0.52	398	0.37	-0.56
401	0.41	-0.37	424	0.43	-0.27
195	0.46	-0.15	66	0.44	-0.24
324	0.70	0.83	407	0.70	0.87
377	0.79	1.34	430	0.82	1.54

Case number 195 and case number 401 whose propensity scores (logits) were 0.46 (-0.15) and 0.41 (-0.37)were matched to control number 66 and control number 424 whose propensity scores equal to 0.44 (-0.24) and 0.43 (-0.27), respectively. Case 390 matched to control 155 and case 324 that was matched to control 407 were exact matches with regards to individual propensity scores. Exact matches are cases in which the difference of the two propensity scores is 0. Using the propensity scores there are 6 exact matches when approximated to 2 places of decimal (and only one exact match with the logits) on the data. Case 377 and control 430 had the greatest difference on both propensity score equal to 0.3 and logits of the propensity scores equal to 0.2. Some graphs depicting the overlap of propensity scores and their logits between the two groups are presented in Appendix II.

The descriptive statistics of the matched covariates are shown on Table 4.5. Compared to the unmatched data, the means and in some cases proportions of the covariates in the cases and controls are pretty similar; subjects in the two groups with nearly equal propensity scores tend to have similar distributions in the covariates used to estimate propensity. The standardized biases are reduced and the propensity score, Parity and body mass index have no bias at all. The p-values are larger in the matched covariates indicating higher level of acceptance that the means/proportions of the cases and controls are equal. Gestational age at birth, UIC, TT4, TPO, sex and birth-weight still have some substantial amount of bias.

Table 4.5: Group comparisons after matching

Continuous Covariates	**Cases (N = 12)**		**Control Group (N = 12)**		**P-value (t-test)**	**Standardized Difference in %**
	Mean	**SD**	**Mean**	**SD**		
Gestational age at birth (weeks)	34.58	0.67	40.17	2.98	<0.0001	258.82
UIC (µg/l)	130.75	114.38	83.08	95.32	0.279	45.28
Total T4 (nmol/l)	9.92	3.32	9.24	2.86	0.597	21.95
Propensity score	0.32	0.24	0.32	0.25	0.964	0
Logit of the propensity score	-0.98	1.31	-0.96	1.35	0.965	1.50
Ordinal/Nominal Covariates	**N = 12**	**%**	**N= 12**	**%**	**P-value (χ^2- test)**	**Standardized Difference in %**
TPO (-ve)	8	66.7	9	75	1.000	18.41
Sex (Female)	3	25	2	11.7	1.000	20.63
Parity (Primiparous)	6	50	6	50	1.000	0
Birth Weight (Low)	2	16.7	1	8.3	1.000	25.40
Body Mass Index (Underweight)	3	25	3	25	1.000	0

Table 4.6 shows the percentage reduction of the standardized bias before and after matching. All the covariates had large bias before matching and bias was reduced on every covariate afterwards. 50% of the covariates had a reduction of over 99%, 25% had over 70% bias reduction, 25% had over 50% and 25% had less than 40% with gestational age having only 19.4%.

Table 4.6: Reduction in Bias

Covariate	**Standardized bias before matching (%)**	**Standardized bias after matching (%)**	**Bias reduction (%)**
Gestational age at birth (weeks)	321.11	258.82	19.39
UIC (µg/l)	72.93	45.28	37.91
Total T4 (nmol/l)	90.75	21.95	75.81
Propensity score	160.86	0	100
Logit of the propensity score	227.75	1.50	99.34
TPO (-ve)	62.33	18.41	70.44
Sex (Female)	63.06	20.63	67.30
Parity (Primiparous)	60.25	0	100
Birth Weight (Low)	54.44	25.40	53.29
Body Mass Index (Underweight)	58.31	0	100

In line with Gu and Rosenbaum (1993) that calipers of 0.2 to 0.6 standard deviations have been shown to remove close to 90% of the bias associated with the matching variable, our caliper of 0.3 has removed 100% of the bias associated with our propensity scores and 99% of that associated with our logits.

CHAPTER 5

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 SUMMARY

The dataset contained 216 subjects of whom 13 were identified as incident cases and 203 were the controls. The preterm babies and the unmatched full-/post-term babies had significantly different baseline values for 10 characteristics (see Table 4.3). After matching, only one covariate remained statistically significant between the two groups and it is in order because gestational age can never be the same for the two groups. The most important thing to report is that bias has been reduced in varying quantities.

We were able to match 12 (92.3%) out of the 13 pre-term babies to comparison subjects in the term-/post-term group. The matched groups were seen to be comparable and a balance of 9 out of 10 background covariates was attained, which is useful for the researcher before he/she decides to commit resources to collecting outcome data. The fit of the propensity model was also adequate. Even though, "when used for matching or stratification, the literature is clear that the main purpose of the propensity score estimation is not to predict treatment based on the baseline covariates but to balance all covariates, therefore, all determinates of the success of a propensity score model are based on the ability of the propensity score to balance covariates." (Caliendo and Kopeinig, 2005; Cavuto *et al*, 2006; Rubin, 2004; Augurzky and Schmidt, 2001)

Brookhart's work further reduces the importance of model fit by suggesting that standard model-building tools designed to create good predictive models of the exposure will not always lead to optimal PS models."(Brookhart *et al*, 2006)

5.2 CONCLUSIONS

The main objective of the study is to use Propensity Score Matching to reduce the self-selection bias involving the non-randomized nature of the control group. Propensity score matching is one way of reducing the selection bias. Propensity scores are an underutilized method for controlling for group differences. They are calculated easily using logistic regression procedures. They provide not only a way of adjusting for group differences but also a diagnostic tool that shows the extent to which treatment groups are comparable.This study showed that the Propensity Score Matching was successful in reducing the bias on the covariates.

Matching on the propensity score is much better than dividing each covariate into two categories and then matching exactly on the categories.

The use of propensity score matching is useful when adjusting for a large number of risk factors and small number of events per variable.

Time and money are saved (when outcome measures have not already been collected or more need to be collected).

Propensity scoring is limited to less than 10 events per variable (EPV), estimated regression coefficients may be biases and SE's may be incorrect (Peduzzi *et al*, 1996; Harrell *et al*, 1985). The biases can be due to inexact matching and it is often better to match all treated subjects and then follow up with analytical adjustments for residuals in the covariates.

Propensity scores can only adjust for observed covariates and work better in larger samples to attain distributional balance of observed covariates and including irrelevant covariates in propensity model may reduce efficiency.

The propensity score complements model-based procedures and is not a substitute for them.

5.3 RECOMMENDATIONS

The logistic regression is just one method of obtaining propensity scores. Another method may be the use of probit model or discriminant model to predict the likelihood of being in the treatment group. The use of the different matching algorithms such as Caliper Matching, Mahalanobis Metric Matching, Kernel Matching, and Stratified Matching are also alternatives to the matching procedure.

5.4 CONTRIBUTION TO KNOWLEDGE

- Through Propensity Score Matching we were able to find a case-control sample that is balanced on baseline covariates and comparable. That is one in which more precise effect of being born preterm on intellectual capability can be assessed.
- From the propensity model, it can be deduced that each extra unit of UIC decreases the odds for having a preterm baby by a multiplicative factor of 0.991 (see Exp (B) column on the UIC row), and each extra unit of TT4 decreases it by 0.731 all other things being equal.
- TPO –ve mothers are 0.164 times less likely to give birth to preterm babies.

- The propensity estimation model suggests that babies of low birth-weight are more likely to be born preterm.

5.5 SUGGESTED AREAS FOR FURTHER RESEARCH

Future researchers should focus on developing and evaluating practical strategies or formal methods to select variables for inclusion in a PS model with an aim of decreasing both the bias and the variance of an estimated treatment effect.

REFERENCES

Ankarali B. C., Sumbuloglu V., Yazici A. C., Yalug I, Selekler M. (2009). *Comparison of different matching methods in observational studies and sensitivity analysis: The relation between depression and STAI-2 scores. Expert Systems with Application ;* **36**:1876-1884.

Augurzky, B. and Schmidt, C. M. (2001). T*he Propensity Score: A Means to an End.* DP No.271. IZA - Institute for the Study of Labor, Boon, Germany. PP. 104.

Austin, P. C. (2008). Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. *Pharmacoepidemiol Drug Saf*; 17(12):1218- 1225.

Austin, P. C. and Mamdani, M. M. (2006). A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med* ; 25(12):2084-2106.

Austin, P. C., Grootendorst, P. and Anderson, G.M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*; 26(4):734-753.

Bergstralh, E.J., Kosanke, J. L. and Jacobsen, S. J. (1996).Software for optimal matching in observational studies. *Epidemiology*; 7(3):331-332.

Berk, R. A. and Newton, P. J. (1985). "Does arrest really deter wife battery? An effort to replicate the findings of the Minneapolis Spouse Abuse Experiment", *American Sociological Review,* **50**, 253-262.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J. and Sturmer, T. (2006). Variable selection for propensity score models. *Am J Epidemiol*; 163(12):1149-1156.

Bryson, A., Dorsett, R. and Purdon, S. (2002). The Use of Propensity Score Matching in the Evaluation of Labour Market Policies. Working Paper 4. Policy Studies Institute and National Centre for Social Research.

Caliendo, M. and Kopeinig S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*; 22(1):31-72.

Campbell, M. J. (2001). *Statistics at Square Two: Understanding Modern Statistical Applications in Medicine*. BMJ Book, PP. 37–55.

Cavuto, S., Bravi, F., Grassi, M. C. and Apolone, G. (2006). Propensity Score for the Analysis of Observational Data: An Introduction and an Illustrative Example. *Drug Development Research*; **67**:208-216.

Cochran, W. G. (1965). "The planning of observational studies of human populations", *Journal of the Royal Statistical Society*, Series A, **128**, 234-255.

Concato, J., Shah, N. and Horwitz, R.I. (2000). Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med*; 342(25):1887-1892.

D'Agostino, R. B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*; 17(19):2265-2281.

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

Dehejia, R. and Wahba, S. (2002).Propensity Score Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics 2002*; 84(1):151-161.

Fiebach, N. H., Cook, E. F., Lee, T. H., Brand, D. A., Rouan, G. W., Weisberg, M. and Goldman, L. (1990). "Outcomes in patients with myocardial infarction who are initially admitted to stepdown units: data from the multicenter chest pain study", *American Journal of Medicine*, **89**, 15-20.

Gu, X. and Rosenbaum, P.R. (1993).Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics*; 2(4):405-420.

Harrell F. E., Jr.,Lee K. L .and Mark D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*; 15(4):361-387.

Harrell F. E., Lee K. L., Mark D. B. and Reichart T. A. (1985). Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treatment Reports*: **69**: 1071- 1077.

Heckman J., Smith J. A .and Clements N. (1997). Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *Review of Economic Studies*; 64(4):487-537.

Hovi P., Andersson S., Eriksson J. G., *et al.* (2007): Glucose regulation in young adults with very low birth weight. *NEJM* 356:2053-2063.

Joffe, M. M., and Rosenbaum, P. R. (1999). Propensity Scores. *American Journal of Epidemiology*, **150**, 327–333.

Limperopoulos C., Bassan H., Sullivan N. R., Soul J. S., *et al*. (2008). Positive screening for autism in ex-preterm infants: Prevalence and risk factors. *Pediatrics* 121(4):758-765.

Mann, C. J. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J*; **20**: 54 – 60.

Månsson *et al*, (2007). On the Estimation and use of Propensity Scores in Case-Control and Case-Cohort Studies. AJE; **166**: 332 – 339.

Martens E. P., Pestman W. R., de B. A., Belitser S. V. and Klungel O. H .(2006). Instrumental variables: application and limitations. *Epidemiology*; 17(3):260-267.

Ming K. and Rosenbaum P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics*; 56(1):118-124.

Muller, J. E., Turi, Z. G., Stone, P. H., Rude, R. E., Raabe, D. S., Jaffe, A. S., Gold, H. K., Gustafson, N., Poole, W. K., Passamani, E., Smith, T. W., Braunwald, E. and The MILIS Study Group (1986). "Digoxin therapy and mortality after myocardial infarction: experience in the MILIS Study", *New England Journal of Medicine*, 314, 265-271.

Peduzzi P., Concato J., Kemper E., Holford T. R. and Feinstein A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*; **49**: 1373-1379.

Perkins S. M., Tu W., Underhill M. G., Zhou X. H., Murray M. D. (2000). The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Saf*; 9(2):93-101.

Roseman, L. (1994). "Using regression and subclassiffication on the propensity score to control bias in observational studies", unpublished report, Harvard University.

Rosenbaum P. R. (1989). Optimal Matching for Observational Studies. *Journal of the American Statistical Association*; 84(408):1024-1032.

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

Rosenbaum, P. R. and Rubin, D. B. (1984). "Reducing bias in observational studies using subclassification on the propensity score", *Journal of the American Statistical Association,* **79**, 516-524.

Rosenbaum P. R. and Rubin D. B. (1985a). The bias due to incomplete matching. *Biometrics*; 41(1):103-116.

Rosenbaum P. R. and Rubin D. B. (1985b). Constructing a Control Group using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*; 39(1):33-38.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*; 127(8 Pt 2):757-763.

Rubin, D. B. (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiol Drug Saf*; 13(12):855-857.

Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*; 52(1):249-264.

Rudner, L. M. and Peyton, J. (2006). Consider Propensity Scores to Compare Treatments. *Practical Assessment Research & Evaluation*, 11(9). Available online: http://pareonline.net/getvn.asp?v=11&n=9

Schendel D. and Bhasin T. K. (2008). Birth weight and gestational age characteristics of children with autism, including a comparison with other developmental disabilities. *Pediatrics,* Jun; 121(6): 1155 – 64.

Shah B. R., Laupacis A., Hux J. E. and Austin P. C. (2005). Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*; 58(6):550-559.

Smith J. A., Todd P. E. Does matching overcome LaLonde's critique of nonexperimental

estimators? Journal of Econometrics 2005; 125:305-323.

Stone, R. A., Obrosky, S., Singer, D. E., Kapoor, W. N. and Fine, M. J. (1995). "Propensity score adjustment for pretreatment differences between hospitalized and ambulatory patients with community-acquired pneumonia", *Medical Care*, 33, AS56-AS66.

Sturmer T., Joshi M., Glynn R. J., Avorn J., Rothman K. J. and Schneeweiss S. (2006). A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*; 59(5):437-447.

Swinscow, T.D.V. (1996). *Statistics at Square One*, 9th edn. (revised by MJ Campbell). London: BMJ Books.

Zhao, Z. (2005).*Sensitivity of Propensity Score Methods to the Specifications*. Institute for the Study of Labor (IZA),Bonn, Germany.

APPENDIX I

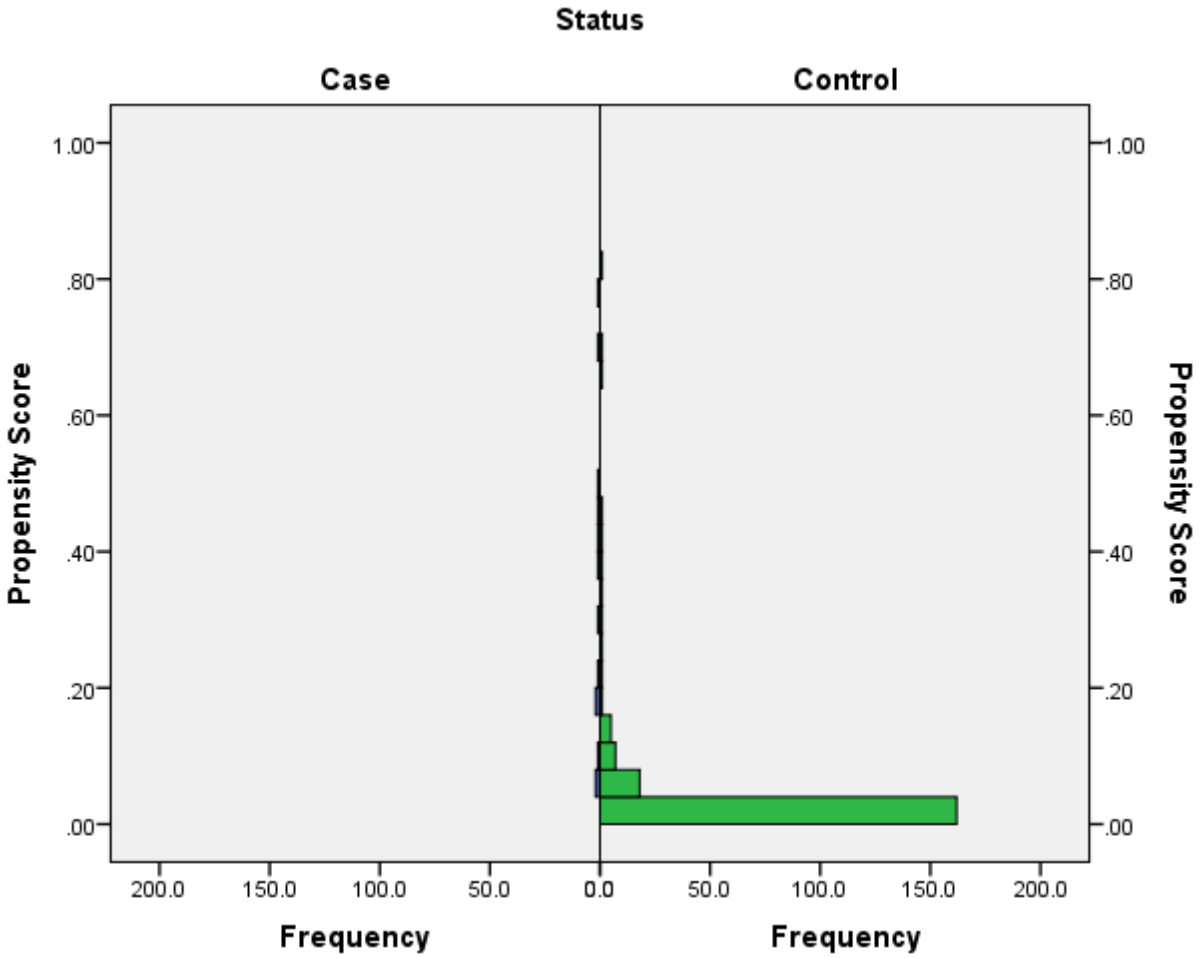
List of potential matches for the 13 cases

S/N	Case	Propensity score	Logit of the PS	Potential matches in control group	Propensity score	Logit of the PS
1	377	0.793	1.344	430	0.824	1.542
2	324	0.696	0.829	407	0.704	0.866
				124	0.651	0.622
3	328	0.488	-0.049	-	-	-
4	195	0.461	-0.155	66	0.442	-0.235
				424	0.433	-0.269
5	401	0.408	-0.374	424	0.433	-0.269
				398	0.365	-0.555
				384	0.330	-0.707
				289	0.320	-0.755
6	286	0.372	-0.524	66	0.442	-0.235
				424	0.433	-0.269
				398	0.365	-0.555
				384	0.330	-0.707
				289	0.320	-0.755
7	415	0.307	-0.814	398	0.365	-0.555
				384	0.330	-0.707
				289	0.320	-0.755
8	42	0.231	-1.204	203	0.244	-1.133
				103	0.207	-1.343
				363	0.181	-1.508
9	405	0.181	-1.511	103	0.207	-1.343
				363	0.181	-1.508
				333	0.155	-1.697
				331	0.149	-1.740
				413	0.144	-1.782
10	411	0.161	-1.652	103	0.207	-1.343
				363	0.181	-1.508
				333	0.155	-1.697
				331	0.149	-1.740
				413	0.144	-1.782
				301	0.125	-1.947
				14	0.124	-1.952

S/N	Case	Propensity score	Logit of the PS	Potential matches in control group	Propensity score	Logit of the PS
11	223	0.115	-2.046	333	0.155	-1.697
				331	0.149	-1.740
				413	0.144	-1.782
				301	0.125	-1.947
				14	0.124	-1.952
				260	0.111	-2.076
				423	0.110	-2.093
				349	0.107	-2.121
				157	0.107	-2.123
				431	0.096	-2.244
				388	0.092	-2.288
12	326	0.064	-2.688	291	0.081	-2.435
				281	0.075	-2.518
				256	0.074	-2.521
				386	0.071	-2.566
				425	0.070	-2.591
				183	0.069	-2.606
				71	0.069	-2.607
				421	0.067	-2.634
				381	0.062	-2.709
				427	0.062	-2.715
				393	0.059	-2.767
				399	0.055	-2.845
				402	0.055	-2.850
				153	0.048	-2.983
				155	0.048	-2.990
13	390	0.047	-3.017	421	0.067	-2.634
				381	0.062	-2.709
				427	0.062	-2.715
				393	0.059	-2.767
				399	0.055	-2.845
				402	0.055	-2.850
				153	0.048	-2.983
				155	0.048	-2.990
				395	0.044	-3.077
				106	0.044	-3.081
				294	0.043	-3.112
				36	0.040	-3.171
				194	0.040	-3.182
				426	0.040	-3.188
				380	0.039	-3.199
				164	0.038	-3.220
				127	0.038	-3.223
				51	0.037	-3.254
				277	0.037	-3.271
				154	0.034	-3.334
				403	0.033	-3.371

APPENDIX II

Figure 1: The Distribution of Propensity Scores for the Case and Control Groups before Matching



- If those in the case group do not overlap (here in terms of the propensity scores or the logits of the propensity scores) with those in the control group, then there is nothing to compare and no type of modelling can help develop information out of thin air. If there is overlap but it is minimal, the information available to infer any treatment effect will lie almost entirely in the few subjects that overlap.

Figure 2: The Distribution of the Logit of the Propensity Scores for the Case and Control Groups before Matching

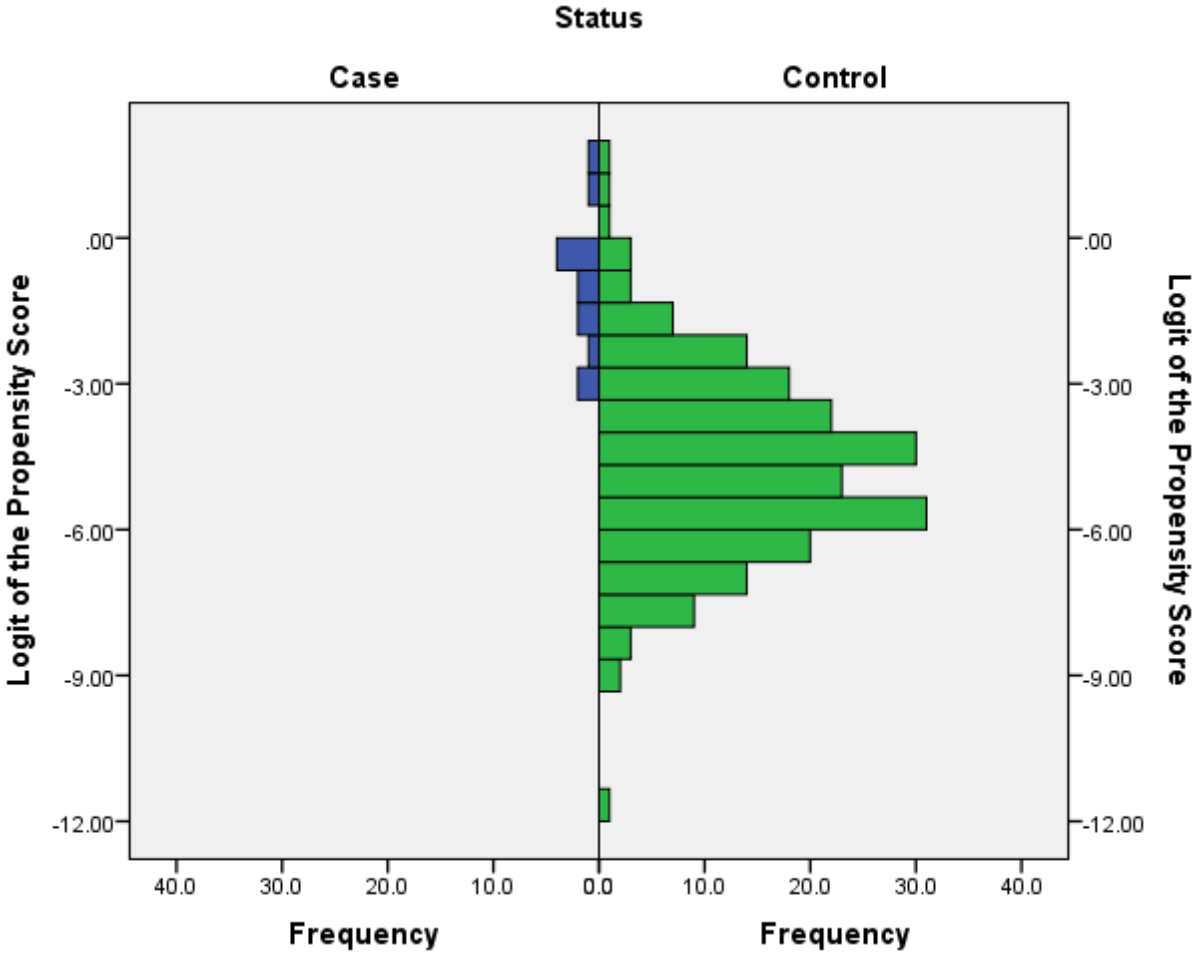
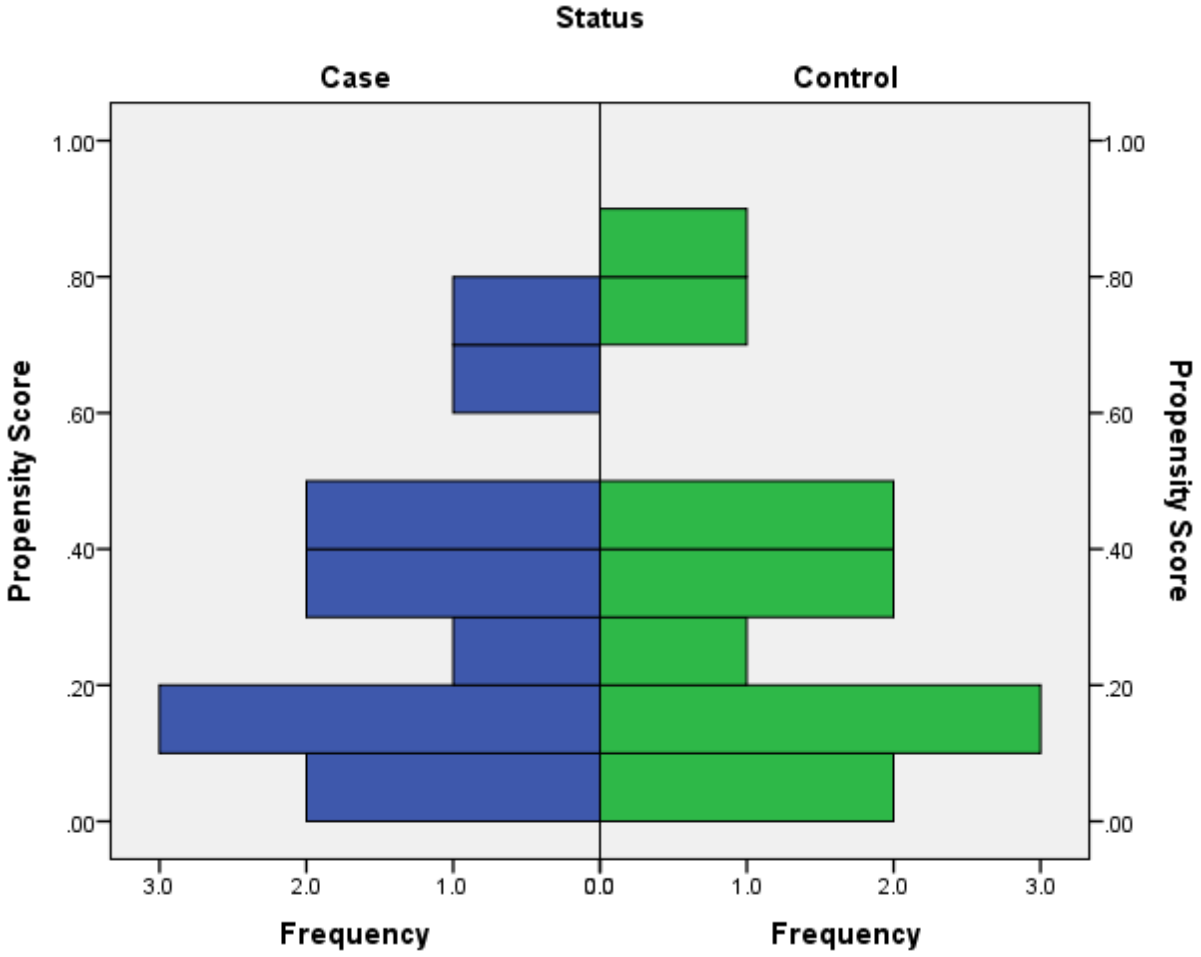
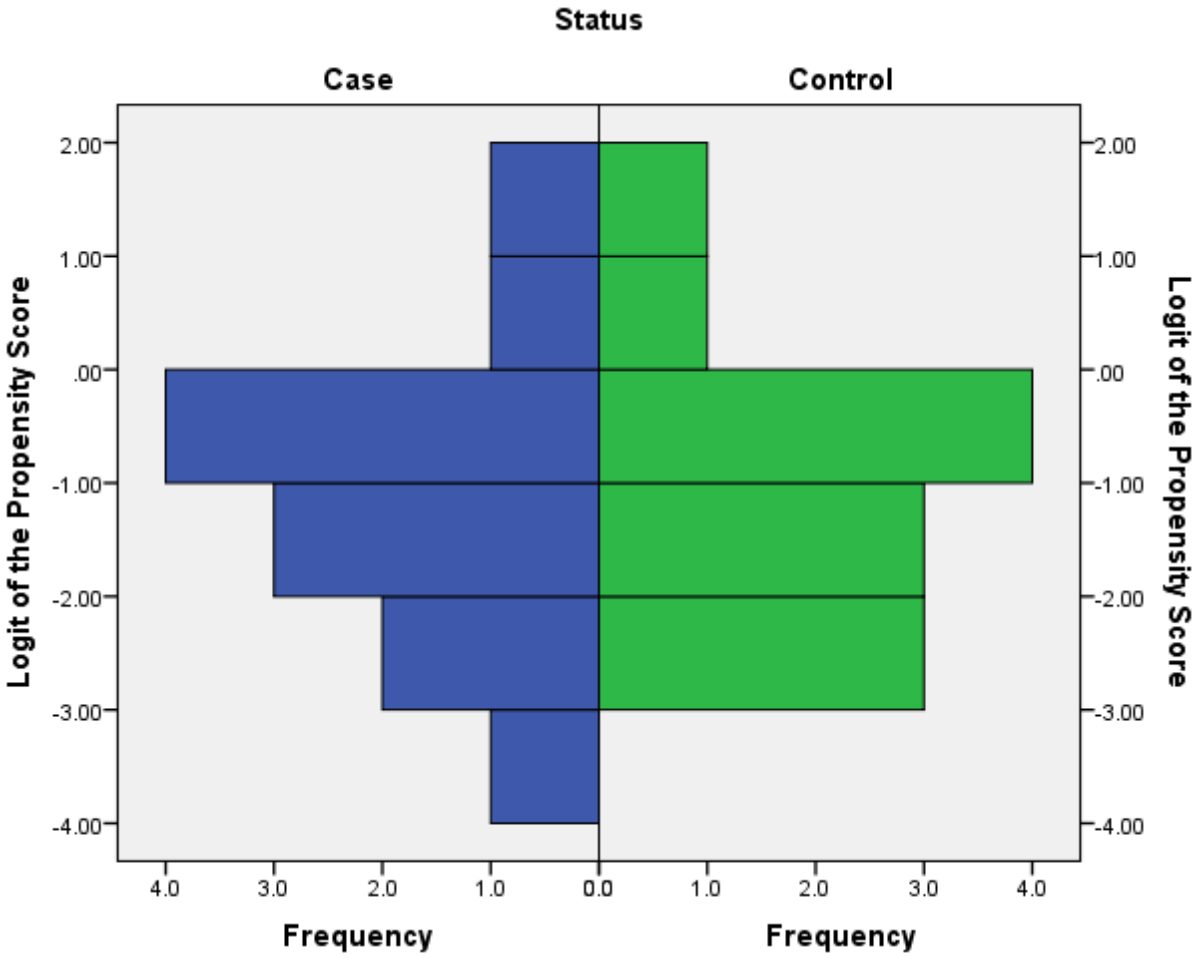


Figure 3: The Distribution of Propensity Scores for the Case and Control Groups after Matching



- With the matched data, the case and control groups overlap almost entirely in terms of the propensity scores and their logits so that we can safely infer causal effects of treatment (preterm birth).

Figure 4: The Distribution of the Logit of the Propensity Scores for the Case and Control Groups after Matching



APPENDIX III

Study Dataset

S/No.	ID	TPO	UIC	GAD	Sex	TT4	BMI	Status	BWT	Parity
1	381	-	78	41	F	9.389	2	0	2	2
2	275	-	470	43	M	12.621	2	0	2	1
3	29	-	408	44	M	12.216	2	0	2	2
4	418	-	366	39	F	11.561	2	0	2	2
5	81	-	22	40	M	11.721	2	0	2	2
6	409	+	260	37	F	23.694	2	0	2	2
7	403	-	98	42	F	10.954	2	0	2	2
8	427	-	52	40	M	11.059	1	0	2	2
9	390	-	140	33	F	8.671	2	1	2	2
10	411	-	20	35	M	10.550	2	1	2	1
11	300	-	257	41	F	16.111	2	0	2	1
12	280	-	68	38	M	11.361	2	0	2	2
13	426	-	22	39	F	12.456	2	0	2	2
14	425	-	18	38	M	9.156	2	0	2	2
15	407	+	78	37	F	10.640	1	0	2	1
16	324	+	90	34	F	3.539	2	1	2	2
17	37	-	133	39	F	19.032	2	0	2	2
18	415	-	207	35	F	6.687	1	1	2	1
19	190	-	257	43	M	7.122	2	0	2	2
20	157	-	36	44	F	8.671	2	0	2	2
21	349	-	52	43	M	6.722	2	0	2	2
22	12	+	620	37	M	8.611	2	0	2	2
23	168	-	370	43	F	10.612	2	0	2	1
24	118	-	265	37	M	8.331	2	0	2	2
25	363	+	91	37	M	9.474	2	0	2	2
26	286	+	17	35	M	12.810	2	1	2	1
27	194	-	48	39	M	14.669	2	0	2	1
28	288	-	346	40	F	12.312	2	0	2	2
29	195	+	30	34	F	12.777	2	1	2	1
30	179	-	256	39	F	8.622	2	0	2	2
31	189	-	347	40	F	16.886	2	0	2	2
32	424	-	38	38	F	9.580	1	0	2	1
33	155	-	52	41	F	110	2	0	2	2
34	28	-	502	36	F	13.868	2	0	2	2
35	69	-	144	38	F	13.834	2	0	2	2
36	322	-	260	40	F	9.885	2	0	2	2
37	136	-	54	39	F	13.071	2	0	2	2
38	100	-	200	38	M	16.873	2	0	2	2
39	419	-	478	38	F	13.285	2	0	2	2
40	417	-	256	42	M	8.932	2	0	2	2
41	164	-	48	44	M	10.342	2	0	2	2
42	103	-	13	44	M	9.755	2	0	2	1
43	38	-	311	40	F	19.662	2	0	2	2
44	326	-	90	35	M	7.492	2	1	2	2
45	246	-	38	44	F	15.861	2	0	2	2
46	32	-	96	36	M	13.475	2	0	2	2
47	399	+	128	40	F	14.231	2	0	2	2

48	235	-	370	38	M	19.622	2	0	2	2
49	196	-	134	43	M	13.081	2	0	2	2
50	203	-	48	43	F	9.626	2	0	2	1
51	83	-	311	40	M	8.261	2	0	2	2
52	154	-	24	40	F	12.867	2	0	2	2
53	57	-	503	41	F	10.822	2	0	2	2
54	106	-	47	40	M	9.926	2	0	2	2
55	278	+	90	39	M	20.681	2	0	2	2
56	244	-	80	40	M	16.822	2	0	2	2
57	153	-	311	42	F	8.322	2	0	2	1
58	215	-	27	41	F	14.853	2	0	2	2
59	176	-	311	40	F	14.611	2	0	2	2
60	26	-	96	43	F	11.583	2	0	2	2
61	391	-	133	40	F	10.876	2	0	2	2
62	130	-	144	40	M	18.011	2	0	2	1
63	416	-	200	45	M	10.275	2	0	2	2
64	389	-	311	42	F	11.547	2	0	2	2
65	402	+	186	43	M	11.155	2	0	2	2
66	184	-	264	40	F	12.602	2	0	2	2
67	23	-	80	37	F	12.761	2	0	2	2
68	223	-	430	35	F	16.228	2	1	1	2
69	227	-	206	42	F	9.821	2	0	2	2
70	80	-	68	39	M	14.763	2	0	2	2
71	71	-	94	44	F	8.626	2	0	2	2
72	423	+	210	36	F	12.022	1	0	2	2
73	422	-	56	43	M	11.130	2	0	2	2
74	408	-	311	39	M	9.910	2	0	2	2
75	283	-	60	43	M	15.060	2	0	2	2
76	220	-	500	36	F	19.328	2	0	2	1
77	36	-	24	40	F	12.346	2	0	2	2
78	412	-	422	39	F	14.284	2	0	2	2
79	156	+	200	36	M	14.318	2	0	2	2
80	243	-	72	39	F	12.671	2	0	2	2
81	137	-	500	41	F	11.226	2	0	2	1
82	431	-	52	40	M	11.563	2	0	2	1
83	383	-	510	41	F	11.867	2	0	2	2
84	14	-	144	42	F	9.611	2	0	2	1
85	432	-	257	43	M	10.727	2	0	2	2
86	265	-	89	41	M	18.490	2	0	2	1
87	291	-	98	38	M	10.911	2	0	2	1
88	259	-	422	37	M	13.220	2	0	2	1
89	193	-	255	39	F	12.601	2	0	2	1
90	229	+	90	37	M	18.011	2	0	2	2
91	84	-	261	39	F	10.621	2	0	2	2
92	293	-	370	42	M	13.286	2	0	2	1
93	17	-	52	42	M	14.889	2	0	2	2
94	147	-	640	38	M	13.285	2	0	2	1
95	292	-	255	42	F	8.322	2	0	2	2
96	413	-	26	41	F	7.855	2	0	2	2
97	202	-	150	40	M	15.854	2	0	2	1
98	298	-	311	42	F	11.750	2	0	2	1
99	333	-	32	45	F	11.868	2	0	2	1
100	310	-	255	37	M	18.431	2	0	2	1

101	214	-	22	38	M	16.977	2	0	2	2
102	274	-	240	37	M	15.691	2	0	2	2
103	211	-	44	44	F	25.183	2	0	2	2
104	201	-	256	37	F	14.680	2	0	2	2
105	65	-	460	37	M	13.962	2	0	2	2
106	401	+	138	35	F	10.516	2	1	2	1
107	86	-	200	44	F	10.328	2	0	2	2
108	295	+	460	39	F	15.612	2	0	2	2
109	116	-	144	39	F	15.355	2	0	2	2
110	405	-	62	35	F	8.440	1	1	2	2
111	146	-	320	41	M	15.187	2	0	2	1
112	414	-	256	38	F	10.976	2	0	2	2
113	334	-	400	41	M	12.452	2	0	2	1
114	392	-	480	38	M	12.515	1	0	2	2
115	35	-	145	39	M	11.641	2	0	2	2
116	205	-	303	42	M	15.607	2	0	2	1
117	222	-	422	40	M	27.881	2	0	2	2
118	145	-	500	39	M	9.321	2	0	2	1
119	384	-	80	39	M	5.887	2	0	2	1
120	191	-	390	41	F	13.795	2	0	2	2
121	331	-	22	39	M	6.328	2	0	2	2
122	267	-	500	40	F	10.640	2	0	2	2
123	430	+	30	36	F	2.909	2	0	2	2
124	228	-	310	43	F	14.313	2	0	2	2
125	356	-	311	41	M	11.484	2	0	2	2
126	169	-	502	45	M	14.773	2	0	2	2
127	279	-	311	39	M	10.348	2	0	2	2
128	332	-	320	41	F	12.796	2	0	2	2
129	68	-	311	38	M	14.661	2	0	2	1
130	421	-	174	36	F	10.967	2	0	2	1
131	117	-	256	41	F	13.821	2	0	2	2
132	289	-	18	44	F	11.681	1	0	2	1
133	388	-	43	40	F	13.452	2	0	2	1
134	55	-	384	45	F	14.315	2	0	2	2
135	249	-	311	38	F	15.224	2	0	2	2
136	192	-	250	37	M	14.087	2	0	2	2
137	200	-	33	37	M	16.281	2	0	2	1
138	135	-	150	43	M	10.391	2	0	2	2
139	226	-	34	39	F	18.128	2	0	2	1
140	242	-	32	36	F	13.241	2	0	2	2
141	348	-	144	42	M	11.703	2	0	2	2
142	387	-	144	37	F	12.745	2	0	2	2
143	287	-	312	39	F	13.101	2	0	2	1
144	104	-	242	38	M	11.808	2	0	2	2
145	398	-	154	38	F	4.875	2	0	2	1
146	210	-	422	42	F	14.210	2	0	2	1
147	125	-	420	36	M	8.621	2	0	2	2
148	42	-	145	34	F	9.633	1	1	2	1
149	301	-	46	37	F	12.281	2	0	2	1
150	294	-	36	43	M	10.326	2	0	2	2
151	96	-	144	42	F	12.837	2	0	2	2
152	231	-	256	40	M	8.101	2	0	2	2
153	385	-	260	45	F	9.383	2	0	2	2

154	277	-	40	39	F	12.226	2	0	2	2
155	178	-	528	41	M	12.181	2	0	2	1
156	321	-	70	42	M	11.699	2	0	2	2
157	127	-	46	43	M	14.855	2	0	2	1
158	345	-	260	43	M	13.235	2	0	2	2
159	393	-	18	42	F	11.221	2	0	2	2
160	441	-	160	42	M	16.681	2	0	2	2
161	382	-	311	43	M	12.021	2	0	2	2
162	166	-	144	38	M	11.226	2	0	2	2
163	386	-	200	39	F	10.035	2	0	2	1
164	74	-	522	37	M	11.861	2	0	2	2
165	307	-	200	42	M	14.816	2	0	2	2
166	221	-	79	44	M	15.862	2	0	2	2
167	248	-	18	42	M	14.811	2	0	2	2
168	404	-	156	37	M	10.682	2	0	2	2
169	206	-	310	41	F	14.560	2	0	2	1
170	241	-	420	42	F	16.821	2	0	2	1
171	344	-	200	40	F	10.739	2	0	2	2
172	225	-	240	39	M	11.624	2	0	2	2
173	144	-	420	40	F	14.221	2	0	2	2
174	380	-	150	42	F	8.981	2	0	2	2
175	197	-	50	39	F	17.344	2	0	2	2
176	218	-	418	44	M	7.665	2	0	2	2
177	428	-	90	40	M	12.608	2	0	2	2
178	314	-	456	37	M	8.301	2	0	2	2
179	284	-	150	44	M	14.821	2	0	2	2
180	281	-	240	39	M	7.282	2	0	2	1
181	163	-	74	40	M	19.213	2	0	2	2
182	66	-	360	42	F	12.364	2	0	1	2
183	260	-	18	40	M	7.510	2	0	2	2
184	397	-	240	37	M	14.921	2	0	2	2
185	299	-	255	42	M	9.899	2	0	2	2
186	111	-	370	40	F	11.775	2	0	2	2
187	257	-	592	40	F	12.270	2	0	2	2
188	395	-	26	42	F	11.990	2	0	2	2
189	165	-	150	41	M	16.347	2	0	2	2
190	346	-	200	41	M	10.351	2	0	2	2
191	377	-	200	35	F	11.711	2	1	1	2
192	261	+	370	38	M	9.690	2	0	2	2
193	13	+	660	44	M	12.321	2	0	2	2
194	59	-	200	43	F	12.681	2	0	2	2
195	73	-	422	38	F	10.391	2	0	2	1
196	328	-	24	35	F	6.824	2	1	2	1
197	15	-	256	36	M	13.768	2	0	2	2
198	161	-	200	39	M	12.081	2	0	2	2
199	90	-	420	38	M	11.603	2	0	2	1
200	120	-	480	40	M	17.223	2	0	2	2
201	320	-	367	45	M	8.992	2	0	2	1
202	143	-	320	40	M	12.676	2	0	2	2
203	126	-	370	38	F	7.112	2	0	2	2
204	51	-	66	42	F	11.459	2	0	2	2
205	72	-	430	38	M	16.803	2	0	2	1
206	266	-	89	43	M	14.160	2	0	2	2

207	313	-	400	40	F	14.290	2	0	2	1
208	258	-	250	43	M	16.370	2	0	2	2
209	256	-	25	41	F	14.690	2	0	2	1
210	124	-	480	43	F	13.224	1	0	1	1
211	262	-	170	41	M	15.000	2	0	2	2
212	99	-	30	40	F	14.322	2	0	2	2
213	27	-	480	39	M	13.115	1	0	2	2
214	2	-	90	42	M	19.861	2	0	2	2
215	347	-	422	37	M	14.778	2	0	2	2
216	183	+	24	39	M	14.819	2	0	2	2