Application of Propensity Score Adjustment to Reduce the Survivorship Bias in Accessing Neurocognitive and Functional Outcomes of Children after Early Cardiac Surgery for Congenital Heart Disease

by

Sara Amiri

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Epidemiology

School of Public Health

University of Alberta

Abstract

Title

Adjusting for survivorship bias for neurocognitive and functional outcome using propensity score and k-means clustering method

Description

In recent years, the number of congenital heart disease survivors has increased with cardiac detection and surgery and perioperative care improvements. However, some survivors may experience poor neurocognitive and functional outcomes after complex cardiac surgery.

It could be related to underestimating survivor outcomes due to a lack of practical statistical methods. We proposed an analytical approach to adjust for the severity of illness, using pre-operative and intra-operative differences among children to develop a propensity score and k-means clustering.

The propensity score and k-means clustering were used to assess the impact of the severity of illness confounders and compare them with crude results.

The analysis included n=235 children with single ventricle congenital heart disease registered at age ≤ 6 weeks in the Western Canadian Complex Pediatric Therapies Follow-up Program between 1997 and 2016.

Preoperative, intraoperative, and postoperative variables were collected. The severity of illness propensity score was calculated based on selected variables. Then, a logistic regression model was set up accordingly.

Neurocognitive and functional outcomes' linear time trends showed that 4.5 years after surgery, FSIQ scores stayed the same, VMI scores increased, and ABAS scores decreased over time in high-risk children.

Using propensity score adjustment helps clarify the actual trend of neurocognitive and functional outcomes in babies with complex conditions.

Table of contents

List of Tables

List of Figures

List of Abbreviations

CHD - congenital heart disease

SV - Single ventricle

MRI - Magnetic resonance imaging

IQ - Intelligence quotient

QOL - Quality of life

CPTFP- Complex pediatric therapies follow-up program

RCT - Randomized control trial

PS - Propensity score

KMC - K-means clustering

US - Ultrasonography

DHCA - Deep hypothermic circulatory arrest

CPB - Cardiopulmonary bypass

LOS - Length of stay

PSM - Propensity score matching

IPW - Inverse probability weighting

PSC - Propensity score as a covariate

SCH - Stollery Children’s Hospital

GRH - Glenrose rehabilitation hospital

WPPSI - The Wechsler Preschool and primary scales of Intelligence

FSIQ - Full-scale intelligence IQ

PIQ -Performance intelligence quotient

VIQ - Verbal intelligence quotient

Beery VMI - The Beery-Buktenica developmental test of visual-motor integration

VMI - Visual motor intelligence

ABAS - Adaptive behavioral assessment system

GAC - General adaptive composite

SD - Standard deviation

MCID - Minimal clinically important difference

ECMO - Extra corporal membrane oxygenation

MV - Mechanical ventilation

ICU - Intensive unit care

ROC -Receiver operating characteristics

AUC – Area under the curve

SES - Socioeconomic status

MEL - Maternal education level

AIC - Akaike information criterion

BIC - Bayesian information criterion

HWSE - Healthy Worker Survivor effect

EPV - Event per variable

GPS - Generalized propensity score

CI - Confidence interval

Chapter 1: Introduction

1.1. Overview

Congenital heart disease (CHD) is the most frequent congenital malformation related to high perinatal, long-term morbidity and mortality (1).

It has been identified as one of Canada's most prevalent and challenging anomalies. According to the perinatal health report 2002, 1 in 100-150 newborns was diagnosed with CHD in Canada, representing significant prevalence (2).

CHD includes structural abnormalities of the heart that develop before birth while the fetus grows in the uterus during pregnancy.

When the heart's chambers, walls, valves, or the blood vessels close to the heart don't develop normally before birth, the result is CHD (3,4).

Single ventricle (SV) condition is considered a type of CHD that might happen when one of the two heart ventricles is too small or weak to function correctly; in other words, the ventricles may be underdeveloped or missing a valve (5). This condition results in one ventricle pumping to the systemic and pulmonary circulations.

In this situation, several open-heart surgeries must be performed to reconstruct the heart anatomy. Three staged surgeries can palliate this condition, including:

1. The Norwood operation

2. The Glenn or hemi-Fontan operation

3. The Fontan operation

Throughout these palliative surgeries, heart and circulatory systems are redesigned (6).

There have been a noticeable number of advancements in the detection and treatment of CHD over the past 50 years. These advances include the detection of CHD using fetal echocardiography, pulse oximetry, chest X-ray and electrocardiogram, magnetic resonance imaging (MRI), and catheter-based methodologies (7,8).

Thanks to improvements in detection, surgery, and perioperative care, the number of survivors with CHD has increased significantly during the past three decades. According to the comprehensive report published in 2017 by the Heart and Stroke Foundation of Canada, less than 20% of babies born with CHD survived to adulthood sixty years ago, while currently, more than 90% of babies reach maturity (9).

1.2. Statement of the problem

It is acknowledged that some children with CHD are subjected to experiencing difficulties with day-to-day activities and academic performance (10), which may last into adulthood (11,12).

Some babies with CHD have intelligence quotient (IQ) scores lower than population norms (13). Some survivors may have a unique neurobehavioral signature, which includes difficulties with social interaction, language, inattention, impulsive behavior, and executive function (14). Moreover, these survivors' quality of life (QOL) can be negatively impacted (15).

Given the remarkable improvements in diagnostic technologies, surgical management, and postoperative care, considerable attention was paid by researchers to neurocognitive trends among the survivors.

It was expected that with the advent of cutting-edge technology in diagnosing and treating CHD and a significant increase in the number of survivors, the neurocognitive trends would also show improvement.

In contrast, improving trends are not always detected with cardiac surgery advances, not exhibiting neurocognitive outcomes improvements over time among these children.

This controversial and thought-provoking matter has shifted the focus to finding its underlying reasons.

The Western Canadian Complex Pediatric Therapies Follow-up Program (CPTFP) Registry was founded in Canada due to the importance of assessing life-saving treatments' effects on long-term outcomes, especially for children's neurodevelopment.

The CPTFP is a multifaceted program in assessment clinics across western Canadian provinces. It provides various services for survivors of life-threatening diseases and determines their long-term neurodevelopmental, neurocognitive, and functional outcomes.

In this program, research is conducted to determine and distribute results, offering survivors of various illnesses different administrations and services and then auditing them to determine potentially modifiable predictors of outcomes. Neurodevelopmental and childhood school-related outcomes after cardiac surgery are a focal interest of this program.

1.3. Study hypothesis

The hypothesis to explain the problem states that neurocognitive and functional outcome trends must be more accurate due to not deploying an optimal method for appraising them.

An important aspect is that babies who previously did not survive due to the complexity of their condition are now entering the outcomes study, contributing to lower outcome scores in survivors. Consequently, these

survivors may counterbalance improved scores in other survivors and cumulatively show unchanged neurocognitive and functional outcomes over the years. Higher scores are balanced by lower scores, preserving the pattern of unchanged average outcomes over time.

These distinct differences stem from a limited selection of study participants, which poses a bias commonly called survivorship bias. Accordingly, this frame of selection bias can result in overly pessimistic judgments. This bias arises from including participants with a broad spectrum of baseline characteristics (16).

A thorough examination of the data is necessary for effective decision-making. Yet, the severity and extent of illness should be considered because the bias is hinged on them, and the solution for overcoming this may be related to the bias condition.

To address this problematic insight, randomized control trials (RCT) may answer it by minimizing confounder impact and defining a more precise trend in results. However, RCTs may only sometimes be viable in biomedical and public health studies and are impossible in our study project framework.

1.4. Study purposes

Therefore, we suggest a statistical method to deal with this survivorship bias, which may prevent us from identifying patterns in post-surgical

outcomes. Propensity score (PS) analysis and the K-means clustering (KMC) method will be implemented in this regard.

PS analysis has recently received much attention in observational cardiovascular intervention research. We will compare the findings using PS adjustment to account for variables affecting post-surgical neurodevelopmental outcomes with the crude trends.

Our specific arguments in this regard can be summarized as follows:

1. Use the PS method to gauge how sick the survivors were.

2. Evaluate the post-operative outcomes adjusted for PS and weigh them against the crude trend.

1.5. Research question

Does applying the PS adjustment facilitate alleviating the effects of survivorship bias and smooth the appraisal of neurodevelopmental outcomes without biased inference?

1.6. Significance of the study

Lower neurocognitive and functional outcomes may disrupt educational and occupational attainment among survivors. Periodic developmental monitoring, screening, assessment, and surveillance throughout childhood might aid in identifying severe impairments and enable the proper therapies

and instruction to be offered that improve later academic, behavioral, psychosocial, and adaptive functioning (17). Therefore, determining the trend over time in these outcomes is essential for planning counselling, surveillance, and intervention services for these children and their families.

1.7. Thesis organization

After the introduction, Chapter 2 thoroughly reflects recent findings and other studies connected to the current study. In Chapter 3, research techniques and statistical analysis will be covered in full. The outcomes of the statistical analysis will be outlined in Chapter 4. The discussion chapter will elaborate on the study's core from different perspectives. Chapter 6 will provide a conclusion and an outlook for future research.

Chapter 2: Literature review

2.1. Improvement in congenital heart disease treatment

CHD is a significant structural heart disease that affects the heart's anatomy and is either detrimental functionally or has the potential to be so (18).

Thanks to worthwhile achievements in pediatric healthcare over the previous 70 years, the number of CHD patients who survive has stunningly surged; more than 90% of babies born in the early 1990s reached adulthood (19, 20).

The upgrades in diagnostic methods, such as catheter interventions (21), surgical treatment of aortic (22) and atrioventricular septal defects (23), Mustard and Senning atrial corrections (24,25), the Rastelli procedure (26), and several surgical innovations, such as the arterial switch operation (27), have all contributed to the betterment (28).

2.2. The complexity of long-term outcomes in survivors of congenital heart disease

Neurologic impairments are among the most prevalent extracardiac consequences in some children with complicated CHD (29,30).

A body of research has shown that children with SV may have a higher risk of adverse neurodevelopmental outcomes than children with other types of

CHD. For some SV patients, performance in the areas of cognition and the gross motor domain was especially poor (31,32)

Children surviving CHD may have outcomes including mild cognitive delay, difficulties with social interaction, and problems with communication skills. In some patients, poor neurodevelopmental outcomes are expected (33).

2.2.1. Prenatal determinants of brain injury

Brain development in babies with CHD may differ from that of non-CHD infants. Abnormal brain growth may start before birth and be influenced by compromised cerebral blood flow. Numerous studies have shown that brain abnormalities in neonates and fetuses are derived from altered cerebral blood flow.

Fetuses with CHD may have inadequate cerebral oxygen delivery due to circulatory abnormalities. Some have autoregulation of blood flow that increases cerebral perfusion and may affect brain growth (34).

MRI and ultrasonography (US) are safe techniques for assessing brain development before and after birth.

Modern prenatal US imaging technology can observe the mother's uterus to study the fetus's brain during pregnancy (35). Clinically, due to its popularity for being transparent and user-friendly, the US is the primary imaging technique for routine evaluation of the fetal brain during pregnancy (36).

However, maternal weight and fetal positioning or some medical conditions may reduce the quality of US images (37). Therefore, prenatal US may be insufficient without fetal MRI.

Due to MRI's propensity for discriminating delicate tissue and liquids, MRI is the ideal imaging method for the brain compared to the US (38). MRI is discussed as more effective at identifying the fundamental intracranial tissue in fetuses with CHD and more precise at measuring the fetal brains (39).

A recent finding in fetal and postnatal MRI has delineated a high frequency of white matter damage, stroke, and bleeding among children with CHD (40).

Furthermore, congenital structural anomalies in the central nervous system are more common and correlate with complex CHD growth (41).

2.2.2. Surgical-based determinants of brain injury

Long-term neurological impairment is interwoven with several preoperative, intraoperative, and postoperative variables. Marking the modifiable predictors at each stage of care is crucial for enhancing long-term brain function, and it can significantly affect children’s longevity and raise their QOL.

2.2.2.1. Preoperative or patient-specific factors

The impact of these variables on neurodevelopmental outcomes has been the subject of extensive research because there is mounting evidence that perioperative injury and complications are linked to these outcomes (42).

Preoperative white matter abnormalities in babies and older children with CHD are associated with adverse neurocognitive outcomes (43,44).

Concurrently, countless explorations are carried out to distinguish perioperative factors associated with adverse long-term neurological outcomes in these children. For instance, variables such as a genetic disease, low birth weight, and the APOE epsilon2 allele were predictors of a worse neurodevelopmental outcome at one year (45).

2.2.2.2. Intraoperative factors

Evidence reveals that some adverse neurodevelopmental outcomes may be associated with intraoperative factors.

To put it differently, some surgical procedures, on the one hand, exponentially raise the number of survivors, and on the other hand, they may also increase the risk of adverse neurological outcomes, including brain injury.

For instance, the potentially detrimental effects of prolonged deep hypothermic circulatory arrest (DHCA) during heart surgery for newborns

and children have been considerably debated in the literature. It is commonly believed that the probability of adverse neurological effects will increase with more extended, unbroken periods of this method (46,47).

A notable reverse association between children's IQ and cardiopulmonary bypass (CPB) duration was also suggested. In other words, a weak but significant negative correlation existed between IQ score and CPB time (48).

In this regard, some investigations of the neurodevelopmental performance of neonates and infants who underwent cardiac surgery have concentrated on potentially modifiable risk factors within surgery, namely intraoperative management techniques.

While variation in intra-operative support, such as the conduct of bypass, is not an easily modifiable risk factor, many still focus on intraoperative management as one mechanism of brain injury that can be accounted for in adjusted statistical models (49).

2.2.2.3. Postoperative factors

Regarding postoperative variables, it was assessed whether there was a relationship between the length of stay (LOS) in hospitals following babies' heart surgery and cognitive outcomes. Compelling findings supported this hypothesis. Even when perioperative events, perfusion times, and

sociodemographic factors are considered, a longer LOS is associated with lower postoperative cognitive performance (50).

Moreover, it has also been demonstrated that plasma lactate concentrations in neonates after cardiac surgery correlate with neurodevelopmental prognosis. It can fundamentally predict postoperative survival rates and adverse neurological sequelae in survivors (51).

2.2.3. Other influential determinants of brain injury

Reportedly, one of the causes of congenital anomalies like atypical prenatal brain development in fetuses with CHD may involve epigenetic and environmental elements.

Environmental factors, including chemicals like pesticides, maternal metabolic conditions like maternal diabetes, drugs taken by a pregnant woman like retinoids, and infection agents like rubella virus and herpes virus, may cause teratogenicity and brain malformation (52).

Besides, epigenetic factors may change brain development during the prenatal and postnatal periods. For example, the placenta's epigenetic processes effectively mediate the growth of the fetus. Furthermore, changes in placental gene expression and signaling during fetal development can significantly alter the developmental program, particularly concerning brain development (53).

2.3. Survivorship bias

2.3.1. Definition

Observational cohort studies are applied to measure how treatments affect patients and how long it takes to achieve a specific result. Since treatment is regularly initiated at various points during a patient's follow-up, it frequently has a time-variant nature. Studies in medical literature mainly overlook how treatment varies over time.

Unlike participants in an RCT, patients in an observational study decide whether and when to begin treatment. Based on that, patients who live longer have more options for choosing a course of treatment; those who die earlier could not receive it by default. When the time-dependent character of treatment is neglected, a survivor treatment selection bias, or, in short, survivorship bias, may develop, severely threatening the validity of the results of the observational study.

This bias impedes the foresight of the natural association between treatment and outcomes and is troublesome to gauge or control (54).

Concerning this, an erroneous association may result from this epidemiologic situation. Noticeable group variations may exist when patients are not randomly assigned to their treatment, which can indirectly impact study

results, and the conclusions drawn from nonrandomized research may be distorted due to this bias.

2.3.2. Survivorship bias and confounding

From an epidemiological perspective, survivorship bias and confounding are two untangled methodological problems that underestimate or overestimate the genuine relationship between outcomes and interventions in observational studies.

To be more precise, survivorship bias is characterized when non-random, systematic circumstances affect who is enrolled in a study.

Alternatively, confounding happens when the same factors are associated with the treatment and the outcome.

Regarding these definitions, survivorship bias and confounding should be effectively addressed to prevent distorted treatment-outcome associations.

Even though these two terms are distinct, the term survivorship bias is used in the literature to refer to confounding (55).

Because the participant's selection method can trigger the confounding effect, entering confounders into the study may be caused by survivorship bias, and consequently, type I error or false positive can occur when the results are incorrectly attributed to the treatment rather than to the confounding variables (56).

To eliminate survivorship bias, emphasis must be placed on confounding.

In observational research, PS, instrumental variables, and multivariable regression modeling are well-known methods used to account for confounding. The most popular techniques applied to correct confounding are multivariable models employing logistic regression (for binary outcomes), Cox proportional hazards models (for survival time), and linear regression (for continuous outcomes).

Regression models track confounding by estimating and describing each variable's contribution to the outcome while maintaining the constant values of all other variables in the model (57).

2.4. Propensity score (PS)

A "propensity score" is the predicted probability of exposure for a particular individual based on relevant variables. The presence of confounding can threaten the study's validity because patients with specific risk factors for an outcome can be systematically directed towards or away from the actual result. To avoid distorting the results and balancing risk factors, PS methods are an efficient way to control baseline confounding (58).

By condensing and simplifying the distributions of numerous measured confounders into a single score based on the likelihood of receiving treatment, researchers can lessen the possibility of selection bias. This

method offers a balanced group for better comparison and would assist in mitigating selection bias in treatment effect estimation (59).

PS formulates a condition called pseudo-randomized experiments to parrot some features of RCT and make the dataset balanced and comparable, making it easier to address bias and draw accurate conclusions (60).

PS methods are frequently used in cardiovascular clinical research; additionally, the effectiveness of various therapies, including hospital-based consult services and at-home care management programs, has been chiefly evaluated in palliative care using PS (61,62).

2.4.1. Propensity scores different approaches

Four PS approaches are matching, stratification, inverse probability weighting, and PS as the covariate. Each has advantages and unique features, providing circumstances to approximate RCT (63).

2.4.1.1. Propensity score matching (PSM)

Forming matched sets of treated and untreated subjects who have similar PS values is known as PS matching. One-to-one or pair matching is the most popular way to use PSM. In this method, treated and untreated subjects are paired up so that their PS values are similar.

After creating a matching selection, the treatment effect can be calculated by directly comparing the outcomes of treated and untreated subjects in the matched sample (64,65).

2.4.1.2. Stratification of propensity score

This method groups subjects into mutually exclusive subsets according to their estimated PS. The estimated PS is used to rank the subjects. Subsets of the subjects are then created based on previously established thresholds. Typically, subjects are divided into five groups of equal size based on the estimated PS quintiles.

It is noted that bias reduction may be improved with more strata being used. Both treated and untreated subjects will have roughly similar PS values within each stratum. The distribution of measured baseline covariates between treated and untreated subjects within the same stratum will be the same when the PS has been correctly specified (66,67).

2.4.1.3. Inverse probability weighting (IPW)

When analyzing the impact of a treatment or exposure, the statistical technique known as inverse probability weighting (IPW) is used to construct otherwise comparable groups. The IPW procedure employs the complete cohort rather than just a subset of variables to match treated and untreated people. It can deal with a wide variety of confounding factors. Each cohort

member is given a weight based on their chance of experiencing the treatment influence under study. When statistical tests or regression models are run, the confounders' effect is diminished or mainly discarded by applying this weight (68).

2.4.1.4. Propensity score as a covariate (PSC)

The other name for this method is PS covariate adjustment. This method involves regressing the outcome variable on an indicator variable that constitutes the state of the treatment and the estimated PS. The estimated regression coefficient from the fitted regression model is used to calculate the treatment's impact. The impact is expressed as an adjusted odds ratio for a logistic model or an adjusted mean difference for a linear model (69).

2.5. K-means clustering (KMC)

Clustering techniques use raw data to create groups of data points based on shared characteristics. KMC aims to divide N observations into K clusters. Each statement belongs to the cluster with the closest mean, the cluster centroid or cluster center, acting as the cluster's prototype. The data space is stratified into Voronoi cells (70).

The KMC process identifies groupings in the data that are implicitly tagged. This method can be implemented to find undiscovered groups in large, complex data sets or to support assumptions about the kinds of groups

present. Any additional data may be readily assigned to the appropriate group once the algorithm has been run and the groups have been formed (71).

Chapter 3: Methods

3.1. Study setting

Two hundred sixty-six babies born with CHD and having cardiac surgery at the Stollery Children's Hospital (SCH) in Edmonton, Alberta, Canada, between 1997 and 2016 were included in this prospective follow-up cohort study. They underwent surgery at SCH at six weeks of age or less. When they were 4.5 years old, their neurocognitive outcomes were assessed at a developmental clinic in their province of birth and for Northern Alberta children at the Glenrose Rehabilitation Hospital (GRH), where the CPTFP is located.

GRH is in Edmonton, Alberta, Canada, as one of the Complex Pediatric Therapies Developmental Assessment Clinics and is part of the inter-provincial Western Canadian CPTFP. In collaboration with pediatric divisions at SCH and the University of Alberta, this medical center provides developmental, functional, language, and cognitive assessments after life-saving surgeries like complex cardiac surgery, chronic renal dialysis, and solid organ transplants.

These programs aim to assess the child's abilities and provide help if needed.

Services are offered to help young children reach early developmental milestones and help preschool children with school placement and learning.

This research was conducted in collaboration with a group of experts from the Department of Pediatrics at the University of Alberta, working in SCH or GRH, including pediatric cardiologists, intensivists, and psychologists.

The compiled dataset for the study had a registration with the CPTFP, consisting of prospectively recorded numerous variables, including the patient's baseline characteristics and surgery-related components and their outcome information.

3.2. Study design and participants

To start, we inspected all the participants' presence in the study. We found that the advent of cutting-edge medical techniques paved the way for all babies to enter the research and have the chance to be treated, regardless of the extent of their illnesses.

Considering all the aspects, we tabled the idea that the association between advancement in cardiac surgery over time and neurocognitive outcomes may be confounded due to surviving participants with different severities of illness. Hence, the severity of heart disease must be effectively controlled to deal with survivorship bias.

The complexity of the disease is a time-varying variable and acts as a confounding factor in the study.

Therefore, we confined our attention to lifting the survivorship bias by evaluating the PS's adjustment impact on determining the severity of the babies' illnesses. And then dichotomize the babies based on the severity of the disease to examine the neurocognitive outcomes in each stratum.

The study participants were two hundred sixty-six infants born with SV at fewer than six weeks old. These high-risk babies underwent Norwood surgery conducted under CPB.

Initially, 22 children were excluded from the study because they had chromosomal abnormalities, of whom died due to the severity of their condition.

The rationale for this exclusion is that multiple studies show that babies with chromosomal abnormalities often have lower neurocognitive scores. Therefore, their inclusion could affect the results, and because of the small number, the adjustment would be inappropriate (72).

Additionally, two babies were lost to follow-up.

Furthermore, seven children did not undergo the neurocognitive assessment for many reasons, including the child who attended the GRH clinic but was

too tired or ill on the day to do the testing or the parents could not stay for the duration of testing.

So, altogether, 31 cases were excluded from the study.

Therefore, the study's final population was 235 individuals without chromosomal abnormalities who underwent Norwood surgery with CPB. For a flowchart of the inclusion/exclusion criteria, please see Figure 1.

3.3 Children's neurocognitive assessment

The following tests were applied to the children to determine outcomes:

3.3.1. The Wechsler preschool and primary scales of intelligence, third edition (WPPSI-III)

An established, individually administered, norm-referenced test of intelligence for young children aged four to seven years and three months is the WPPSI-III. The test embodies seven subtests, and the findings all reveal a full-scale intelligence quotient (FSIQ) containing a performance intelligence quotient (PIQ) and a verbal intelligence quotient (VIQ) (73).

Block design, information, matrix reasoning, vocabulary, picture concepts, word reasoning, coding, symbol search, receptive vocabulary, and picture naming were among the WPPSI-III subtests administered (74).

3.3.2. The Beery-Buktenica developmental test of visual-motor integration, fifth edition (Beery VMI-V)

Especially in school-based practice, this exam determines eligibility for occupational therapy services. However, visual motor integration (VMI) is used not only as a screening or assessment tool but also as an outcome measure to determine improvements in visual-motor integration skills after handwriting interventions. The VMI is intended to examine a person's ability to integrate visual and motor skills and to provide appropriate treatment (75).

The Beery VMI investigates the participants' capability to combine their visual and motor skills by having them replicate progressively complicated geometric patterns. Children between the ages of two and eight can complete the short and full-format tests by copying drawings of geometric shapes placed in increasing order of complexity (76).

3.3.3. The Adaptive behavior assessment system second edition (ABAS-II)

A thorough norm-referenced rating scale, the ABAS-II, checks adaptive behavior and skills in people from birth to 89 years. Based on the participant's age bracket, there are three different rating forms:

1. For children ages 0 to 5, a parent/primary caregiver form and a teacher/daycare provider form.

2. A parent form and a teacher form for children ages 5 to 21

3. Adult forms are all available for ages 16–89.

Regarding gender, race/ethnicity, and parental education, the standardization sample is sizable (>4,000) and representative of United States data from 1999 to 2000.

The ABAS-II includes ten skills comprising communication, community use, functional pre-academics, home living, health and safety, leisure, self-care, self-direction, social, and motor skills across four domains consisting of conceptual, social, practical, and the general adaptive composite (GAC) score, a complete overall indicator of adaptive abilities. The GAC score was chosen as the target variable to quantify how well children can adapt (77) since children's adaptive functioning (78,79) is the destination of this test.

The GAC compares a person’s global adaptive skills to the adaptive skills of others in the same age group from the standardization sample (80).

Therefore, FSIQ, VMI, and ABAS GAC were our neurocognitive and functional outcomes in this study.

Benchmarks for each metric have a population mean of 100 and a standard deviation (SD) of 15. These test scores are calculated based on the norms of the general population.

From a statistical perspective, an SD of 15 means 68% of the general population has scored between 85 and 115. Also, 95% of the general population has a score within 2 SD, representing 95% of the individuals who have a score between 70 and 130.

Accordingly, data that does not fall between +2 SD and -2 SD from the mean is considered an outlier, occurring in 2.5% of the normative population.

However, from a clinical point of view, we use ½ SD for clinical results interpretation because, based on the minimal clinically important difference (MCID), using the one-half SD represents the minimally important difference in the clinical outcomes. In other words, the smallest change in the results would be identified as clinically significant (81).

Concerning MCID, any changes equal to 7.5 in any of these tests will be considered clinically significant.

To ensure the accuracy and reliability of the test results, the test information was precisely documented, and the calculation of the score results of each

test was double-checked with the skilled pediatric psychologist working in the GRH.

3.4. Statistical analysis for logistic regression

3.4.1. Predictor variables in the PS model building.

The given dataset consisted of many variables presenting each child's baseline characteristics. For a table of the variables, please see Table 1.

To decide whether to keep or discard a variable, we consulted with an expert pediatric cardiologist and intensivist at the first step. The potential predictors were chosen based on the time of measurement, being clinically significant, and not having a missing value, a small sample size, or an unbalanced group.

Therefore, birth year, year of initial surgery, weight less than 2500 grams at surgery, sepsis, and extracorporeal membrane oxygenation (ECMO) were dropped from the list of variables.

The remaining variables, including age at surgery, gestational age, birth weight, weight at surgery, inotrope score, serum lactate, base deficit, arterial pH, total mechanical ventilation (MV) days, CPB time, cross-clamp time, DHCA time, creatinine level, total days chest open, sex, prenatal diagnosis, cardiopulmonary resuscitation, convulsion, and dialysis, were

noted as the variables to investigate their presence in the model. For the table of the variables considered, please see Table 2.

3.4.2. Propensity score mechanism

PS uses a predicted probability of group membership based on observed predictors, often determined through logistic regression.

The estimated PS e(xi) for subject i (i = 1,..., N) is the conditional probability of being assigned to a treatment given covariate xi.

PS is the probability of receiving the treatment of interest (Z = 1), conditional on observed baseline characteristics.

We can calculate a PS conditional on observed pre-operative and intra-operative variables that reflect the severity of illness and, therefore, likely influence post-surgical outcomes.

3.4.3. Determining the outcome for logistic regression

We had to consider an outcome variable among the abovementioned factors to perform PS model building in a logistic regression model. On the one hand, the chosen variable had to serve as a standard for evaluating the complexity and severity of CHD. Additionally, it had to have a significant impact on the neurocognitive outcome.

In the first step, we reviewed many papers to find the best option. In this regard, we found two salient predictors in our dataset that can meet the above criteria: total MV days and intensive care unit (ICU) LOS.

LOS, used for department management, quality assurance, and hospital planning, is the total number of days a patient is hospitalized. Additionally, it might infer the efficiency and efficient use of the hospital's resources (82).

We found a direct connection between illness severity, LOS, and patients' outcomes following ICU discharge. It has been empirically proven that sicker babies with more complex conditions after surgery have extended ICU stays (83).

In addition, some patients who leave the ICU are at high risk of developing persistent cognitive impairment and having low QOL. Worse cognition and QOL scores are linked to poor baseline patient characteristics and ICU events (84).

A study has shown that positive cognitive stimuli are decreased while risk factors for cognitive impairment are multiplied during ICU time intervals (85,86).

Furthermore, some ICU survivors are at risk of having new or deteriorating impairments in various cognitive functions and common psychiatric disorders like depression, anxiety, and post-ICU syndromes (87).

One study showed that hospital design, fear of pain, and the uniformed staff were cited as stress-inducing factors by the patients discharged from the ICU. Furthermore, impairments in memory, executive function, social interactions, and neuropsychiatric morbidities have also been connected to deprived environments, such as hospitals or institutions (88). Psychological stress and potential cognitive difficulties are also linked to MV time, surgical interventions, sedation, and pain medication (89).

Besides, in terms of MV, one of the most crucial clinical variables determining how well pediatric cardiac surgery will turn out is the length of MV. Following cardiac surgery, the MV remains prolonged for multiple reasons—the more complex the baseline baby characteristic, the more extended the MV time (90).

Therefore, total MV days and ICU LOS were the best options for the logistic regression's outcome.

In the second step, we presented our options to the pediatric cardiologist and intensivist experts, who advised us to choose ICU LOS. Therefore, the ICU LOS was regarded as the outcome of the regression.

However, we were also advised to include death by 30 days as another outcome for the regression. Since the PS regression reflected the perioperative severity of illness, early death (which could occur even with short ICU LOS) was considered significant.

Death within 30 days is the standard time surgeons consider for surgical mortality, and this measure is applied to assess, evaluate, and compare hospital quality (91).

We maintained that the ICU LOS for the first surgery or death within 30 days after the first surgery would represent the children's severity of illness and can be used as the outcome for logistic regression to estimate a PS for perioperative seriousness of the disease.

3.4.4. Assigning a cut-point for the logistic regression outcome

We considered all pre- and intra-operative available variables, potentially influencing post-surgical outcomes to start formulating the regression.

ICU LOS was labeled a continuous variable in our dataset, so we had to convert it into a binary variable to run logistic regression. Therefore, we needed a cut-off value.

The best cut-off value for clinical applications can be chosen using receiver operating characteristic (ROC) curves, which graphically plot true positives versus false positives across a range of cut-offs. The trade-off between sensitivity and specificity facilitates determining the best cut-offs.

Maximizing the sum of sensitivity and specificity as a diagnostic threshold criterion is formally equivalent to minimizing the sum of false negative and false positive misclassification likelihoods. In ROC curve analysis, the ideal

limit has the clinically desirable property of maximizing the correct diagnosis rate and minimizing the overall misdiagnosis rate, so ten days were coined as an appropriate threshold for ICU LOS (92,93,94), with the highest accuracy (83.1%), sensitivity (84.9%), and specificity (54.1%), the highest area under the curve (AUC) (73.3%), and clinical relevance.

3.4.5. Model building for propensity scores.

Initially, we started model building by sticking to the defining groups based on the ICU LOS, so code for group 1 as babies who had an ICU LOS longer than ten days or died within 30 days and code for group 0 as babies who stayed in the ICU for fewer than ten days and did not experience death within 30 days.

Then, we tracked the purposeful variable selection method and entered all the predictor variables one by one in the logistic regression to end up with their corresponding p-values and unadjusted odds ratios. Based on the results of this univariate analysis in the unadjusted logistic regression, if the p-value was less than 0.1, we kept the variable in the model. Otherwise, the variable was discarded from the model.

In the next stage, we sought multivariate analysis. All selected variables from the previous univariate analysis were entered into the model simultaneously. Based on the adjusted odds ratio and corresponding p-value among the remaining variables, the ones with a significant p-value less than

0.05 were chosen to stay in the final logistic model. We finalized the adjusted logistic regression and best-fit model equations, including age at surgery, total MV days, DHCA time, and total chest open days.

For our data, simply, the systematic part of the logistic regression model can be written as follows:

$$\text{Log}\frac{P(Zi=1|Xi)}{P(Zi=0|Xi)} = \log\frac{e(Xi)}{1-e(Xi)} = \beta 0+\beta 1\text{Xi}$$

Z represents ICU LOS for more than ten days or death within 30 days after the first surgery. X constitutes a set of age at surgery, total MV days, DHCA time, and total chest open days.

3.4.6. Calculating the propensity score

Then, we placed the values related to each baby in the formula to calculate each patient's predicted probability of Z.

$$\text{Pr }(\text{zi}=1|\text{xi})= \frac{1}{1+e^{-(\beta 0+\beta iXi)}}$$

By plugging each number for each patient into the above formula, we computed a predicted probability for each patient separately. Afterward, as

mentioned above, the calculation results provided us with 235 scores for all the babies.

3.4.7. K-means clustering on the propensity score.

Clustering is an unsupervised statistical procedure where data points are grouped into different sets based on their degree of similarity. In contrast to the supervised system, clustering does not use labeled data. One of the various types of clustering is partitioning clustering, subdivided into K-means clustering and Fuzzy C-means clustering.

In KMC, the features are compared, and all objects having similar characteristics are clustered together.

These methods have been used previously for assigning subjects with a similar distribution of PS to the same group (95).

In Fuzzy C-means clustering, each point has the probability of belonging to more than one cluster, while in KMC, a single object cannot apply to two different groups.

The partitioning clustering procedure has indispensable types of algorithms. One of the famous ones is centroid-based clustering, which organizes the data into non-hierarchical clusters. KMC is the most widely used centroid-based clustering method. This design is efficient but sensitive to primary conditions and outliers (96).

3.4.8. Determining the optimal number of clusters

Therefore, we utilized the K-means center-based clustering method on PS to find infants with a similar distribution of observed covariates. These methods helped us discriminate between babies with severe and non-severe conditions.

In this regard, we emphasized the Silhouette method to find the optimal number of clusters and interpret and validate consistency within data clusters. This process computes Silhouette coefficients for each point that measure how much a topic is similar to its group compared to other sets by providing a concise graphical representation of how well each object has been classified. The Silhouette coefficients were rated for each point, and the average was taken for all the samples to get the Silhouette score.

Silhouette width is generally recognized as an index for assessing the individual's fit in the classification, the quality of clusters, and the entire category (97).

3.4.9. Defining the clusters

The results of KMC stratified babies into two clusters. In each cluster, some babies had ICU LOS of more than ten days, and some had ICU LOS of less than ten days. Then, we traded off the clustering results with the PS outcome variable (ICU LOS or death within 30 days) by making a two-by-

two table. In each cluster, we focused on the babies with the highest proportion. The two concordant pairs of two-by-two tables included the babies with the higher percentage, which facilitated categorizing infants into severe and non-severe groups. Based on the findings, we drew our attention to the severe group allegedly prone to inefficient neurocognitive outcomes.

3.5. Statistical analysis for linear regression

3.5.1. Predictor variables for neurocognitive and functional outcomes

The given dataset had some variables related to neurocognitive outcomes. To decide whether to keep a variable, we separately set up purposeful variable selection through univariate analysis for each outcome: FSIQ, VMI, and ABAS GAC. Variables with a p-value less than 0.1 were selected for the next step. Additionally, we took advantage of the expert cardiologist and intensivist team's and psychologists' ideas, and they confirmed the clinically significant contribution of the following variables:

1. Socioeconomic status (SES)

This factor was assessed using the Blishen index. SES measures an individual's standing in the community. It refers to a person's social status determined by wealth, occupation, and social class. Typically, it focuses on wealth, education, and income.

SES has been associated with functioning across various neurocognitive domains, covering language, memory, executive functioning, and social-emotional processing. Extensive research has documented socioeconomic disparities in academic performance (98).

2. Maternal education level (MEL)

It is readily acknowledged that parents' educational level is a critical predictor of cognitive, language, and motor outcomes in every age bracket of the babies' life span, ranging from preterm infants to older ages. For instance, evidence has shown that as MEL increased from less than high school to university or higher, cognitive and language scores were higher for infants born at 29 weeks gestation (99).

The maturation of the human brain during early development necessitates the coordinated growth of all brain regions over time. Both lower MEL and lower levels of cognitive development were found to be associated with highly coordinated brain volume growth. These variations were most noticeable in older kids (100,101).

3. Total hospital days at SCH after the first admission.

In much of the literature covering cardiac research results, the number of days a baby stays at the hospital was cited as an influential factor for determining future cognitive outcomes.

This factor can act as a marker of the many challenges babies struggle with during hospitalization.

A more extended postoperative stay is associated with worse later cognitive function, even when adjusted for perioperative events, perfusion times, and sociodemographic variables; however, more research is required to figure out the mechanisms underlying this relationship. Postoperative LOS after infant heart surgery may be a simple surrogate for various events resulting in later adverse cognitive outcomes (102).

3.5.2. Assessing collinearity among the selected variables

As we know, the correlation between predictor variables causes them to be unable to predict the dependent variable's value independently. Consequently, the statistical significance of an independent variable will be undermined, so we assessed the collinearity among the named variables. Based on that, the MEL and SES were correlated, so the MEL was discarded from the analysis.

We also included the birth year as a variable that can elaborate on each trend’s changes over time, a significant aim of this study.

Therefore, the final multiple linear regression for each neurocognitive outcome consisted of three variables: SES, all hospital days, and birth years as time.

3.5.3. Different neurocognitive models implementing PS adjustment.

We recruited the calculated PS once in the context of a covariate and once as a weight in the multiple linear regression model to adjust for the baseline severity of perioperative illness (using clinical and patient characteristics) and estimate the neurocognitive outcomes (103,104).

In IPW, fitted probability values were grounded on the participant group number. For group 1, where Z = 1, or the infants with ICU LOS longer than ten days or those who died within 30 days, it was calculated in this way: (1/predicted probabilities).

For group 0, where Z = 0, or for the infants with ICU LOS lower than ten days or who did not die before 30 days, PS was calculated in this way: (1/(1-predicted probability)).

In PSC, the calculated predicted probability will be entered as a covariate in the final multiple linear models to adjust for the baseline clinical and patient characteristics reflecting the perioperative severity of illness.

The multiple linear regression was implemented as all the outcomes were continuous; here is the neurocognitive trend.

To simplify, the systematic part of our multiple linear regression model can be written as follows:

$E[Y|X] = \beta 0 + \beta 1 X1$

Y is the outcome variable (here: FSIQ, VMI, or ABAS GAC), and X constitutes the set of SES, all hospital days at SCH after the first admission and birth year.

The estimation of the effect size or coefficient in the simple linear regression could be obtained from this formula:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2}$$

Likewise, for the IPW simple linear regression, the formula would be as follows:

$$b_1 = \frac{\sum_{i=1}^{n} w_i(x_i - \bar{x})(y_i - \bar{y})}{w_i(x_i - \bar{x})^2}$$

Wi specifically represents the PS calculated from a multiple logistic regression for each participant that influences the coefficient as a weight.

Concerning these formulas, the effect sizes for three different multiple linear regression models, including adjusted, IPW-adjusted, and PSC-adjusted, were calculated, considering 0.05 as the significance level. Although some of the three variables did not achieve the assigned statistical significance to be retained, they were kept in the model due to their clinical significance.

Adjusted multiple linear regression only had SES, all hospital days, and birth year, not including the PS.

While IPW- and PSC-adjusted multiple regression models covered the same list of variables plus the calculated PS.

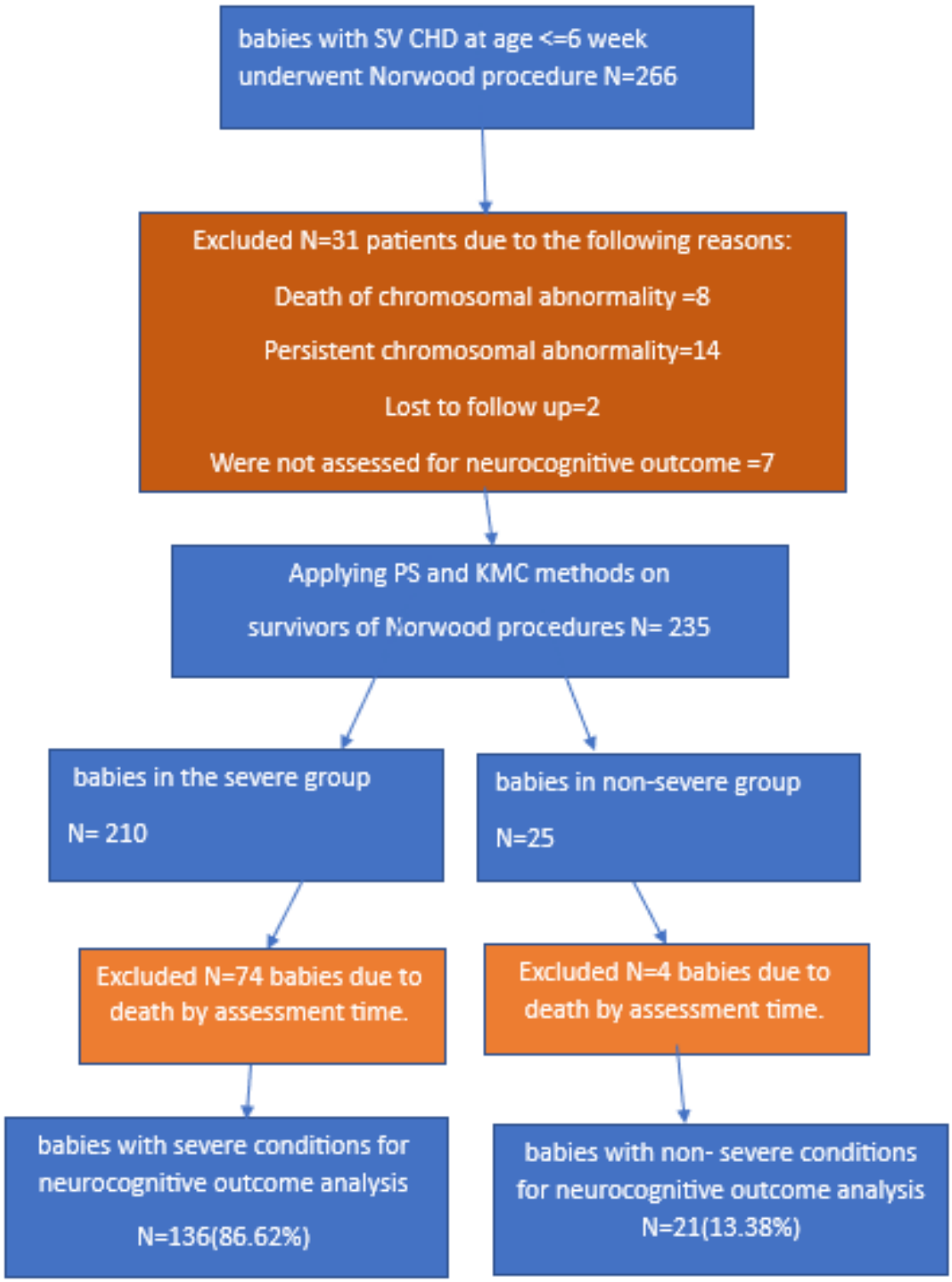
We compared IPW- and PSC-adjusted models with the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as the two ordinary penalized goodness of fit criteria (105). The aim was to find out which approach would better provide results. The model with the smaller AIC and BIC frames was more suitable, and all statistical analysis was performed using R software version 4.0.4.

Chapter 4: Results

4.1. Description of the cohort study

The study group consisted of 266 children with SV admitted at the SCH in Edmonton, Alberta, Canada, between 1997 and 2016 who underwent the Norwood procedure. Figure 1 depicts the levels of screening and population selection.

Figure 1: Flowchart of analysis of cohort.



4.2. Description of the predictor variables in the preoperative and intraoperative stages of palliative surgeries

The fundamental description of the study population is tabulated in Table 1. It incorporated patient and procedure-related features and some continuous and categorical variables for 235 participants. The mean of each variable was reported separately. 20 (8.5%) of the 235 infants weighed less than 2500 grams at surgery, and the average (SD) of the birth weight z-score and age at surgery were, respectively, -0.12 (1.0) and 10.5 (6.9) days. 90 (38.3%) of the children were female.

Table 1: Baseline and clinical characteristics of the study population(n=235) from the palliative surgery

Continuous Variables	N	Mean (SD)
Age at surgery	235	10.49 (6.9)
Gestational age, weeks	235	38.77 (1.7)
Birth weight, Z-score	235	-0.12 (1)
Weight at surgery	235	3.28 (0.6)

Highest mod inotrope score, mcg/kg/min	235	7.7 (20.5)
Highest serum lactate, mmol/L	235	3.42 (2.7)
Lowest base deficit, mmol/L	235	-5.04 (4.5)
Lowest arterial pH	235	7.26 (0.1)
Total ventilation days pre-op	235	6.91 (6.6)
Cardiopulmonary bypass time	235	123.5 (54.3)
Cross-clamp time	235	48.51 (22.4)
Deep hypothermic circulatory arrest (DHCA) time	235	25.06 (17.7)
Day 1 Highest inotrope score, mcg/kg/min	235	15.85 (16)
Day 1 Highest serum lactate, mmol/L	235	7.42 (4)

Day 1 Lactate time to <2 mmol/L, hours	235	19.29 (15.4)
Day 1 Lowest base deficit, mmol/L	235	-2.73 (4.9)
Day 1 Lowest arterial pH	235	7.27 (0.1)
Day 1 Highest creatinine, umol/L	235	61.19 (20.8)
Day 2-5 Highest inotrope score, mcg/kg/min	234	16.28 (36.5)
Day 2-5 Highest serum lactate, mmol/L	234	3.98 (4.1)
Day 2-5 Lowest base deficit, mmol/L	234	-2.26 (4.4)
Day 2-5 Lowest arterial pH	234	7.3 (0.1)
Day 2-5 Highest creatinine, umol/L	234	70.91 (31.9)
Total days chest open	235	6.1 (6.3)

Categorical Variables	N	%
Sex, female	90	38.30%
Birth year		
1997-1998	19	8.10%
1999-2000	23	9.80%
2001-2002	21	8.90%
2003-2004	23	9.80%
2005-2006	22	9.40%
2007-2008	28	11.90%
2009-2010	30	12.80%
2011-2012	25	10.60%
2013-2014	29	12.30%
2015-2016	15	6.40%
Year of initial surgery		
1997-1998	19	8.10%

1999-2000	23	9.80%
2001-2002	20	8.50%
2003-2004	24	10.20%
2005-2006	22	9.40%
2007-2008	28	11.90%
2009-2010	30	12.80%
2011-2012	25	10.60%
2013-2014	29	12.30%
2015-2016	15	6.40%
Weight <2500g at surgery	20	8.50%
Prenatal diagnosis, antenatal diagnosis	149	63.40%
Re- Cardiopulmonary bypass, yes	50	21.30%
Sepsis, positive bc	36	15.30%

Convulsions anytime, yes	23	9.80%
Cardiopulmonary resuscitation anytime, yes	33	14%
Dialysis anytime, yes	65	27.70%
Extracorporeal membrane oxygenation (ECMO), includes ECPR, yes	29	12.30%
Death by 30 days	17	7.20%
Death by 4.5 years	78	33.20%

4.3. Description of the propensity score model

Table 2 splits the variables nominated to assess their presence in the multiple logistic regression. The pre-, intra-, and early post-operative variables were recorded in the CPTFP registry database. For each, the number of participants who spent less or more than ten days in the ICU or died within 30 days of surgery was delineated. The corresponding

unadjusted and adjusted odds ratios and p-values were calculated at each level.

To be more precise, to achieve the final multiple logistic regression, two prominent screening stages were conducted: first, based on the unadjusted odds ratio, we selected the variables with a significant p-value, which was less than 0.1, encompassing age at surgery, the highest mod inotrope score, the lowest base deficit, the lowest arterial PH, total MV days, DHCA time, day 2–5 highest inotrope score, day 2–5 lowest base deficit, day 2–5 highest creatinine, and total days with the chest open.

Secondly, based on the adjusted odds ratio, the variables with a significant p-value less than 0.05 were assigned to be kept in the model, the remaining ones comprising age at surgery, total MV days pre-operative, DHCA time, and whole days with the chest open.

Table 2: Logistic regression for propensity score(n=235).

Variables	ICU LOS <10d and survived within 30 days [N=16, 6.81%]	ICU LOS ≥10d or died within 30 days [N=219, 93.19%]	Unadjusted OR (95%CI)	Adjusted OR (95%CI)
Continuous	Mean (SD)	Mean (SD)		
Age at surgery	13.94 (10.3)	10.24 (6.6)	0.95 (0.9, 1) * P-value= 0.048	0.85 (0.8, 0.9) P-value= 0.002
Gestational age, weeks	38.88 (1.8)	38.76 (1.7)	0.96 (0.7, 1.3) P-value= 0.793	--
Birth weight, Z-score	-0.11 (0.8)	-0.12 (1)	0.99 (0.6, 1.7) P-value= 0.979	--
Weight at surgery	3.23 (0.5)	3.28 (0.6)	1.18 (0.5, 2.9) P-value= 0.718	--

Highest mod inotrope score, mcg/kg/min	1.62 (3.2)	8.14 (21.1)	1.16 (1, 1.4) * P-value= 0.048	NS P-value= 0.657
Highest serum lactate, mmol/L	2.75 (2.1)	3.47 (2.8)	1.17 (0.9, 1.7) P-value= 0.308	--
Lowest base deficit, mmol/L	-1.94 (3.7)	-5.27 (4.5)	0.77 (0.7, 0.9) * P-value= 0.002	NS P-value= 0.272
Lowest arterial pH	7.34 (0.1)	7.25 (0.1)	6.7e-6 (3.3e-9, 4.6e-3) * P-value= 0.001	NS P-value= 0.935
Total ventilation days pre-op	2.56 (3.2)	7.23 (6.7)	1.29 (1.1, 1.5) * P-value= 0.002	1.34 (1.1, 1.7) P-value= 0.004
Cardiopulmonary bypass time	117.81 (45)	123.92 (55)	1 (1, 1) P-value= 0.664	--
Cross-clamp time	47.19 (16.5)	48.6 (22.8)	1 (1, 1) P-value= 0.807	--

Deep hypothermic circulatory arrest (DHCA) time	12.12 (10.7)	26.01 (17.7)	1.07 (1, 1.1) * P-value= 0.004	1.09 (1, 1.2) P-value= 0.004
Day 1 Highest inotrope score, mcg/kg/min	10.69 (6.4)	16.23 (16.4)	1.06 (1, 1.1) P-value= 0.122	--
Day 1 Highest serum lactate, mmol/L	5.89 (1.9)	7.53 (4.1)	1.18 (1, 1.5) P-value= 0.103	--
Day 1 Lactate time to <2 mmol/L, hours	16.53 (8.7)	19.49 (15.8)	1.02 (1, 1.1) P-value= 0.457	--
Day 1 Lowest base deficit, mmol/L	-1.89 (5)	-2.79 (4.9)	0.96 (0.9, 1.1) P-value= 0.475	--
Day 1 Lowest arterial pH	7.27 (0.1)	7.27 (0.1)	1.28 (0, 317.3) P-value= 0.935	--
Day 1 Highest creatinine, umol/L	52.94 (12)	61.79 (21.2)	1.03 (1, 1.1) P-value= 0.1	--
Day 2-5 Highest inotrope score, mcg/kg/min	6.81 (4.6)	16.98 (37.7)	1.15 (1, 1.3) * P-value= 0.017	NS P-value= 0.877

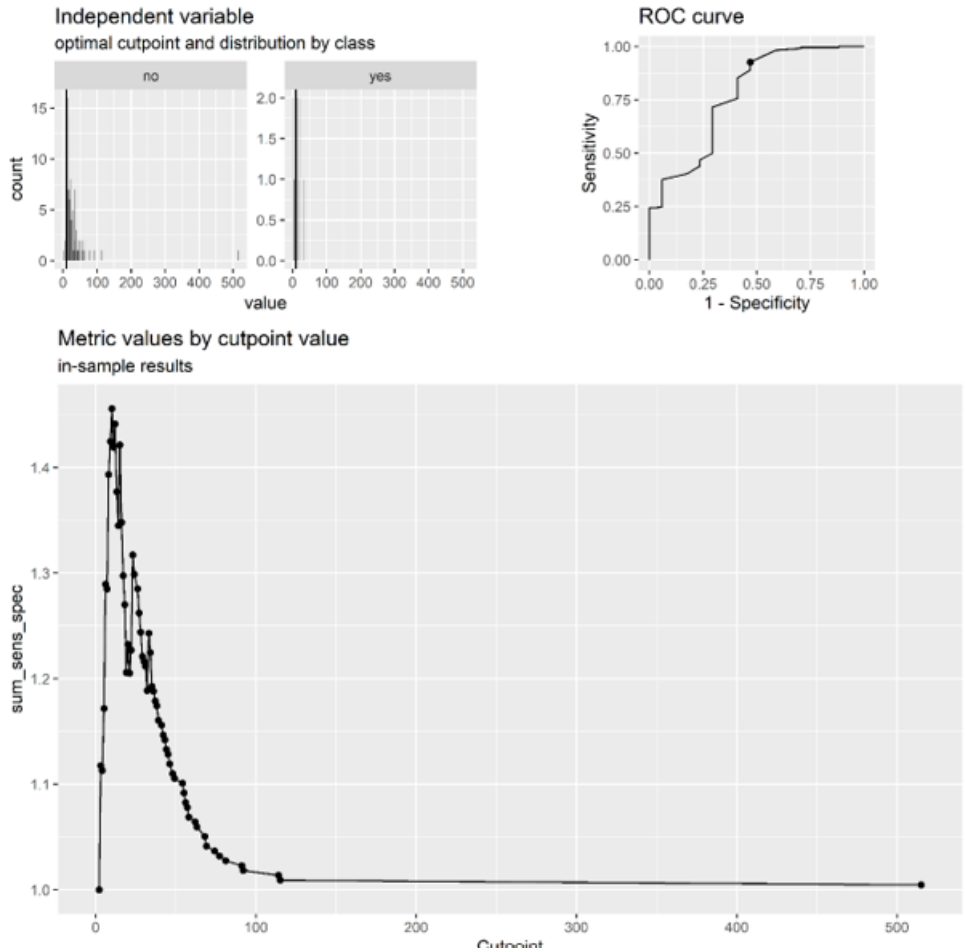
Day 2-5 Highest serum lactate, mmol/L	2.58 (1.2)	4.08 (4.2)	1.33 (1, 2.1) P-value= 0.132	--
Day 2-5 Lowest base deficit, mmol/L	-0.52 (4.4)	-2.39 (4.3)	0.89 (0.8, 1) * P-value= 0.086	NS P-value= 0.278
Day 2-5 Lowest arterial pH	7.32 (0)	7.3 (0.1)	0.01 (0, 11.2) P-value= 0.283	--
Day 2-5 Highest creatinine, umol/L	51.12 (16.1)	72.36 (32.3)	1.04 (1, 1.1) * P-value= 0.011	NS P-value= 0.510
Total days chest open	3.25 (0.9)	6.31 (6.5)	1.42 (1.1, 1.8) * P-value= 0.005	2.02 (1.3, 3.4) P-value= 0.003

Categorical	N (%)	N (%)	OR (95%CI)	
Sex, female	4 (25%)	86 (39.27%)	1.94 (0.7, 7.1) P-value= 0.264	--
Prenatal diagnosis, antenatal diagnosis	10 (62.5%)	139 (63.47%)	1.04 (0.3, 2.9) P-value= 0.938	--

Convulsions anytime, yes	1 (6.25%)	22 (10.05%)	1.68 (0.3, 31) P-value= 0.625	--
Cardiopulmonary resuscitation anytime, yes	1 (6.25%)	32 (14.61%)	2.57 (0.5, 47.2) P-value= 0.369	--
Dialysis anytime, yes	3 (18.75%)	62 (28.31%)	1.71 (0.5, 7.6) P-value= 0.414	--

The results for determining the optimal cut point for ICU LOS based on the sum of specificity and sensitivity are illustrated in Figure 2.

Figure 2: Optimal cut-point for ICU LOS based on the sum of sensitivity and specificity (n=235)



4.4. Description of clustering method on propensity score

Figure 3 resonates with using the unsupervised clustering method through the Silhouette procedure to regulate the optimal number of clusters for the study. Following this method led us to two clusters.

Figure 3: Optimal number of clusters using the Silhouette method for PS (n=235)

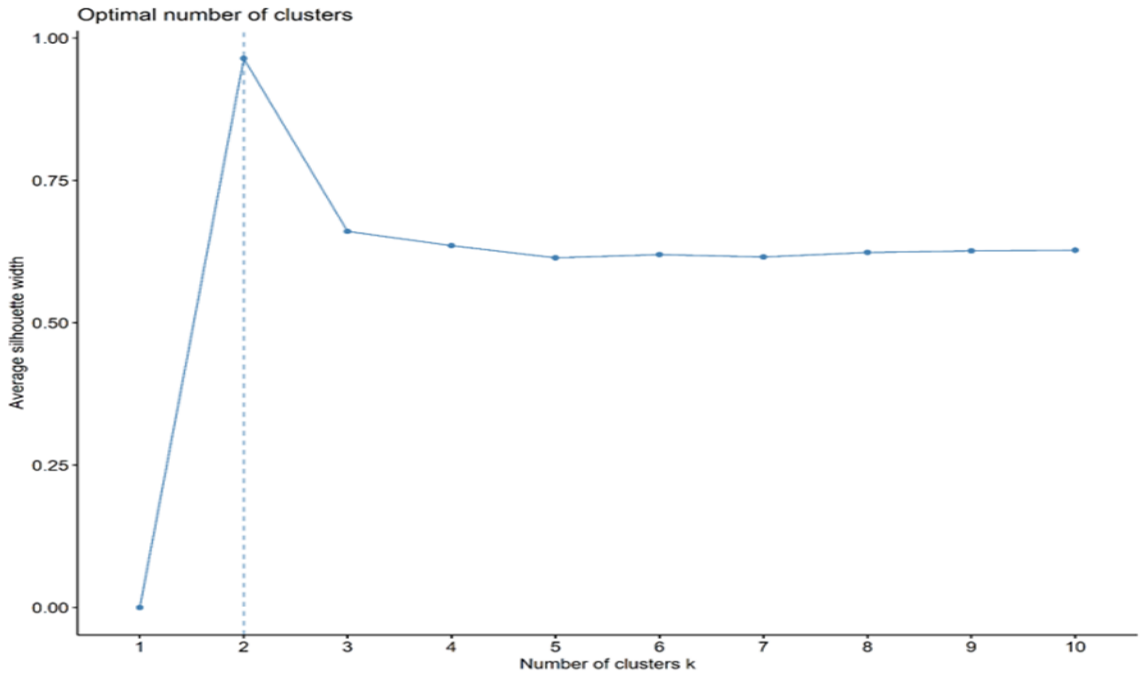


Table 3 reports a two-by-two table based on the clustering results, giving us the insight to label the population under study into the severe condition group and the non-severe one. The two-by-two table based on PS outcome, ICU LOS and early mortality stratified all the participants, marking two

concordant pairs that paved the way for categorizing infants into severe and non-severe groups.

In cluster 1, the most significant proportion, 60%, belongs to babies with less than ten days of ICU stay, while in cluster 2, babies with more than ten days grabbed the 95% proportion.

Indeed, of 235 infants, 210 were dichotomized into babies who struggled with the severity of their disease, and 25 were not confronted with the severe degree of sickness.

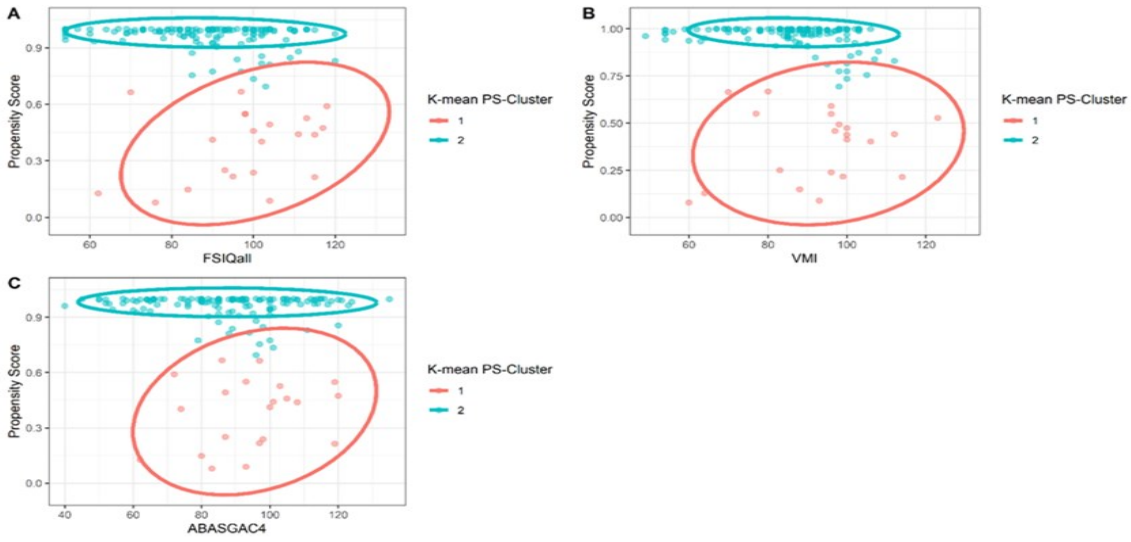
Table 3: Labeling the clusters based on PS to discriminate the severe from the non-severe group (n = 235).

Cluster	ICU LOS <10d and survived within 30 days	ICU LOS >=10d or died within 30 days	Total	Label
1	15 (60%)	10 (40%)	25 (100%)	Not-Severe
2	1 (0.5%)	209 (95.5%)	210 (100%)	Severe
Total	16	219	235	-

After omitting 78 children due to death by 4.5-year assessments, 157 babies remained in the study for the final analysis of neurocognitive outcomes. Of the 157 participants, 21 (13.37%) were classified as non-severe and 136 (86.63%) as severe.

In Figure 4, scatter plots A, B, and C independently visualize the distribution of 157 babies into each cluster for each neurocognitive and functional outcome. As it is represented, there is no overlapping among the clusters.

Figure 4: Descriptive plots for each cluster and each neurocognitive outcome(n=157)



4.5. Description of multiple linear regression for neurocognitive outcome

Table 4 describes the neurocognitive and functional outcomes and their potential predictors for 157 survivors. The average (SD) scores of FSIQ, VMI, and ABAS GAC for the entire cohort were 89.6 (16.1), 87.02 (13.9), and 88.59 (19.0), respectively. The average (SD) scores for the Blishen

Index and all hospital days were 44.49 (13.8) and 40.72 (72.1), respectively.

Table 4: Neurocognitive criteria tested and potential variables (n=157).

Neurocognitive and Functional Outcomes	Mean (SD)/ N (%)
WPPSI-III Full-Scale Intelligence Quotient (FSIQ)	89.6 (16.1)
Beery Visual-Motor Integration (VMI)	87.02 (13.9)
The Adaptive Behavior Assessment System - Second Edition (ABAS-II) General Adaptive Composite (GAC)	88.59 (19.0)
Predictor Variables Considered	
SES(Blishen Index)	44.49 (13.8)

All hospital days (associated with CPB) at SCH after the first admission	40.72 (72.1)

Table 5 features the goodness-of-fit appraisal results for fitted models. As a result, the PSC-adjusted model presented less value in AIC and BIC than the IPW-adjusted model, which means it can show the detailed trend better than the IPW-adjusted model.

Table 5: Goodness of fit criteria for fitted model (n=157)

Outcomes	AIC	BIC
FSIQ		
Multiple Linear Regression-IPW adjusted [Red Line]	1316.44	1331.69
Multiple Linear Regression-PSC adjusted [Green Line]	1302.68	1320.98

VMI		
Multiple Linear Regression-IPW adjusted [Red Line]	1265.87	1281.12
Multiple Linear Regression-PSC adjusted [Green Line]	1261.90	1280.20
ABAS-II GAC		
Multiple Linear Regression-IPW adjusted [Red Line]	1376.22	1391.47
Multiple Linear Regression-PSC adjusted [Green Line]	1359.93	1378.23

AIC: Akaike information criterion, BIC: Bayesian information criterion

Tables 6 and 7 summarised multiple linear regression effect sizes yielded for each outcome. The results for the entire sample and the severe group were embedded in the table.

Table 6: Multiple linear regression effect size for study outcomes based on the entire sample (n=157)

Outcome	Variables	Adjusted Effect Size	IPW[#]- Adjusted Effect Size	PSC[#]-Adjusted Effect Size
FSIQ	Blishen Index	0.15 (-0.03, 0.33) P_Value= 0.105	0.13 (-0.05, 0.32) P_Value= 0.148	0.18 (-0.01, 0.36) P_Value= 0.057
	All Hospital Days	-0.04 (-0.08, -0.01) P_Value= 0.014	-0.04 (-0.08, -0.01) P_Value= 0.012	-0.04 (-0.07, -0.005) P_Value= 0.026
	Birth year (Time)	0.26 (-0.23, 0.76) P_Value= 0.29	0.16 (-0.33, 0.66) P_Value= 0.512	0.01 (-0.52, 0.53) P_Value= 0.988
VMI	Blishen Index	0.1 (-0.06, 0.26) P_Value= 0.224	0.09 (-0.06, 0.25) P_Value= 0.244	0.11 (-0.04, 0.27) P_Value= 0.157

	All Hospital Days	-0.03 (-0.06, 0.003) P_Value= 0.083	-0.03 (-0.05, 0.003) P_Value= 0.08	-0.02 (-0.05, 0.01) P_Value= 0.123
	Birth year (Time)	0.51 (0.08, 0.93) P_Value= 0.02	0.45 (0.03, 0.87) P_Value= 0.037	0.34 (-0.12, 0.8) P_Value= 0.147
ABAS-II GAC	Blishen Index	0.08 (-0.14, 0.3) P_Value= 0.486	0.06 (-0.16, 0.28) P_Value= 0.598	0.10 (-0.12, 0.32) P_Value= 0.361
	All Hospital Days	-0.05 (-0.09, -0.01) P_Value= 0.014	-0.05 (-0.09, -0.01) P_Value= 0.012	-0.05 (-0.09, -0.01) P_Value= 0.023
	Birth year (Time)	-0.31 (-0.89, 0.28) P_Value= 0.300	-0.41 (-1.01, 0.19) P_Value= 0.178	-0.55 (-1.19, 0.08) P_Value= 0.086

Table 7: Multiple linear regression effect size for study outcome based on the severe group (n=136)

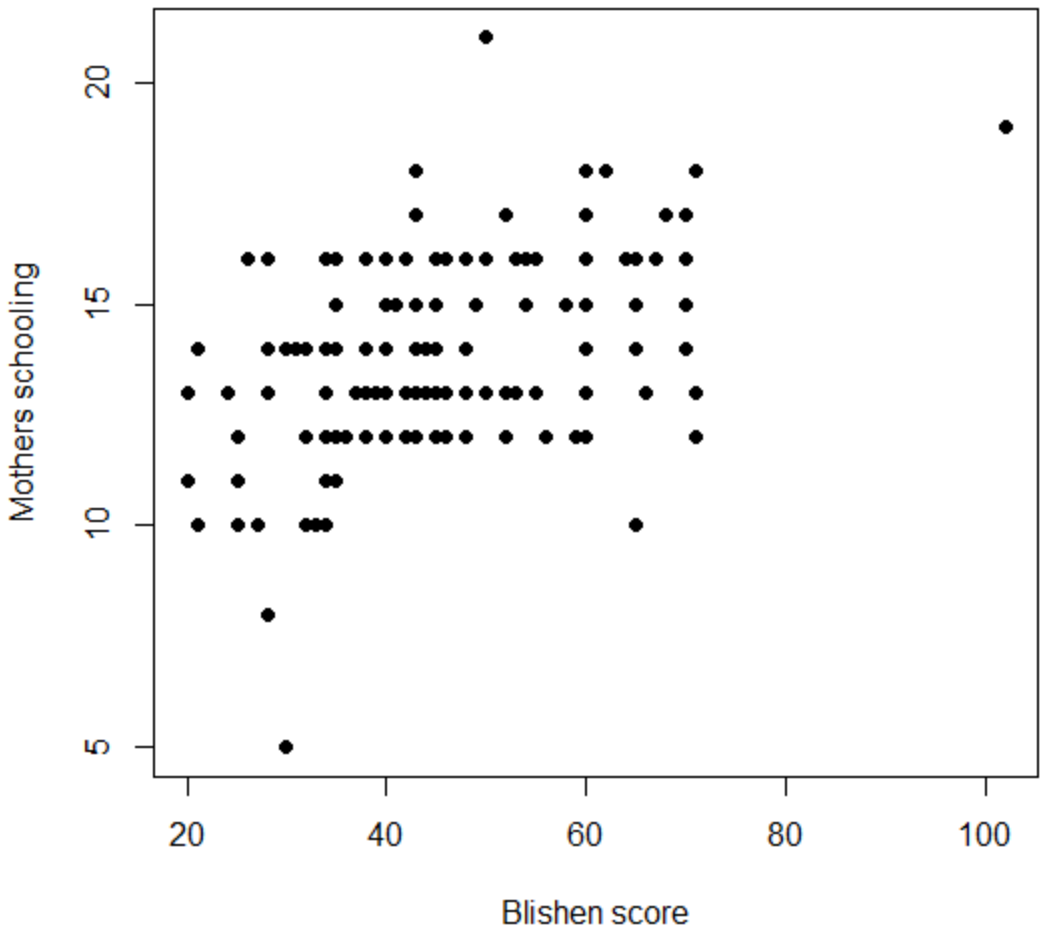
Outcome	Variables	Adjusted Effect Size	IPW[#]- Adjusted Effect Size	PSC[#]-Adjusted Effect Size
FSIQ	Blishen Index	0.15 (-0.05, 0.34) P_Value= 0.137	0.14 (-0.05, 0.34) P_Value= 0.140	0.14 (-0.05, 0.33) P_Value= 0.145
	All Hospital Days	-0.04 (-0.08, -0.01) P_Value= 0.025	-0.04 (-0.08, -0.01) P_Value= 0.024	-0.04 (-0.07, -0.002) P_Value= 0.035
	Birth year (Time)	0.04 (-0.49, 0.57) P_Value= 0.879	0.02 (-0.51, 0.56) P_Value= 0.938	-0.14 (-0.69, 0.41) P_Value= 0.611
VMI	Blishen Index	0.1 (-0.06, 0.26)	0.1 (-0.06, 0.27) P_Value= 0.211	0.09 (-0.07, 0.25) P_Value= 0.247

		P_Value= 0.233		
	All Hospital Days	-0.02 (-0.05, 0.01)	-0.02 (-0.05, 0.01)	-0.02 (-0.05, 0.01)
		P_Value= 0.133	P_Value= 0.141	P_Value= 0.189
	Birth year (Time)	0.41 (-0.04, 0.86)	0.39 (-0.07, 0.84)	0.2 (-0.26, 0.65)
		P_Value= 0.073	P_Value= 0.093	P_Value= 0.398
ABAS-II GAC	Blishen Index	0.06 (-0.17, 0.3)	0.06 (-0.17, 0.3)	0.06 (-0.18, 0.29)
		P_Value= 0.605	P_Value= 0.602	P_Value= 0.629
	All Hospital Days	-0.05 (-0.1, -0.01)	-0.05 (-0.1, -0.01)	-0.05 (-0.09, -0.01)
		P_Value= 0.015	P_Value= 0.016	P_Value= 0.020
	Birth year (Time)	-0.54 (-1.19, 0.1)	-0.55 (-1.2, 0.1)	-0.7 (-1.37, -0.03)
		P_Value= 0.099	P_Value= 0.098	P_Value= 0.042

Table 8 shows that despite the influential effect of MEL on the neurocognitive outcome, it was dropped from the study due to collinearity with SES.

Table 8: Result of collinearity among variables

	SES	MEL
SES	1	
MEL	0.48 P_value<0.001	1



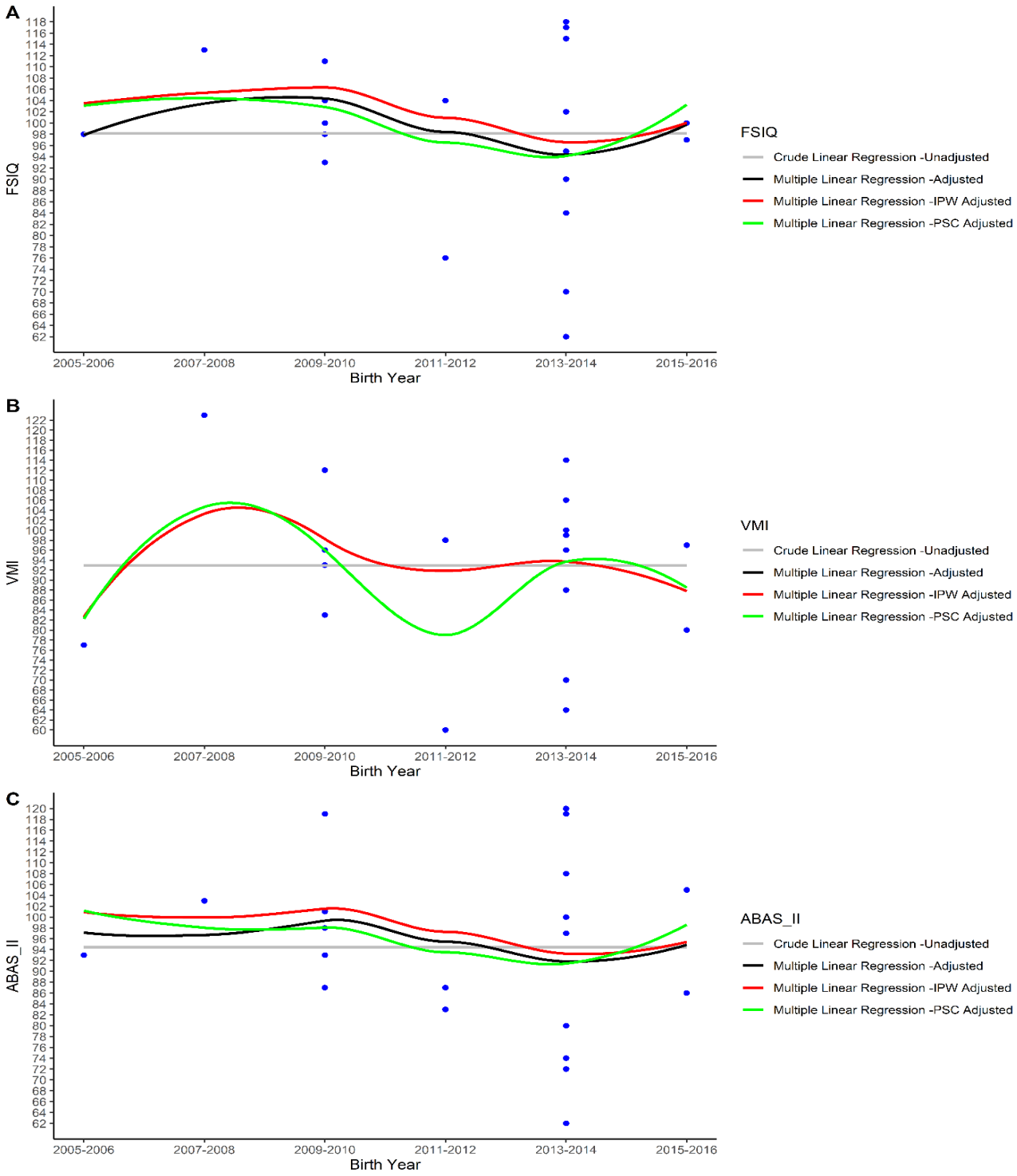
In addition to the various multiple linear regressions defined beforehand to facilitate the comparison for each group of the severity condition, a crude or unadjusted linear regression was designed, containing neither the SES, all hospital days, nor the PS.

Therefore, each plot portrays four lines at the same time.

Figure 5 shows all the neurocognitive and functional trends for the 21 babies in the not-severe group over the birth year; estimates are unstable because of sparse data, including data that started from the 2005-2006 birth year.

In plot B for the VMI score, the black line for multiple linear regression-adjusted and the green line for multiple linear regression-PSC-adjusted overlapped.

Figure 5: Linear trends of outcomes for the not-severe group, A for FSIQ, B for VMI, and C for ABAS GAC



Figures 6, 7, and 8 display the linear trends of neurocognitive and functional outcomes over the birth year based on crude, adjusted, IPW-adjusted, and PSC-adjusted models.

Each figure consists of two plots: plot A, the fitted lines over the entire sample, and Plot B, the matched lines over the severe group, which encountered more challenges.

Figure 6 denotes the FSIQ score trend:

1. FSIQ score trend for the entire sample:

In Plot A on the entire sample, the PSC-adjusted model captures a slight downward trend from 1997–1998 to 2001–2002, followed by a plateau trend to 2005–2006, and then an upward trend until 2011–2012, experiencing a period of stability until 2015–2016. The time trend of this model has the corresponding P-value= 0.988.

In plot A on the entire sample, the IPW-adjusted model portrays a stable trend from 1997–1998 to 2001–2002, followed by a slight increase until 2007–2008 and another steady drift until 2015–2016. P-value = 0.512 for the time trend of this model.

For the time trend, a P-value=0.29 was reported for the multiple linear regression-adjusted models.

2. FSIQ score trend for the severe group:

However, for the severe group in Figure 6 Plot B, the FSIQ score trend in the PSC-adjusted model shows a slight decrease from 1997–1998 to 2005–2006, followed by a slightly upward trend till 2015–2016. The time trend of this model has the corresponding P-value= 0.611.

For the IPW-adjusted model, the trend is very close to a flat line until 2003–2004, followed by a small increase till 2015–2016. The time trend of this model has the corresponding P-value= 0.938.

The multiple linear regression adjusted model for the severe group represented the time trend with the P-value=0.879.

Figure 6: Linear trends of FSIQ score over birth years.

Plot A for the entire sample and plot B for the severe group.

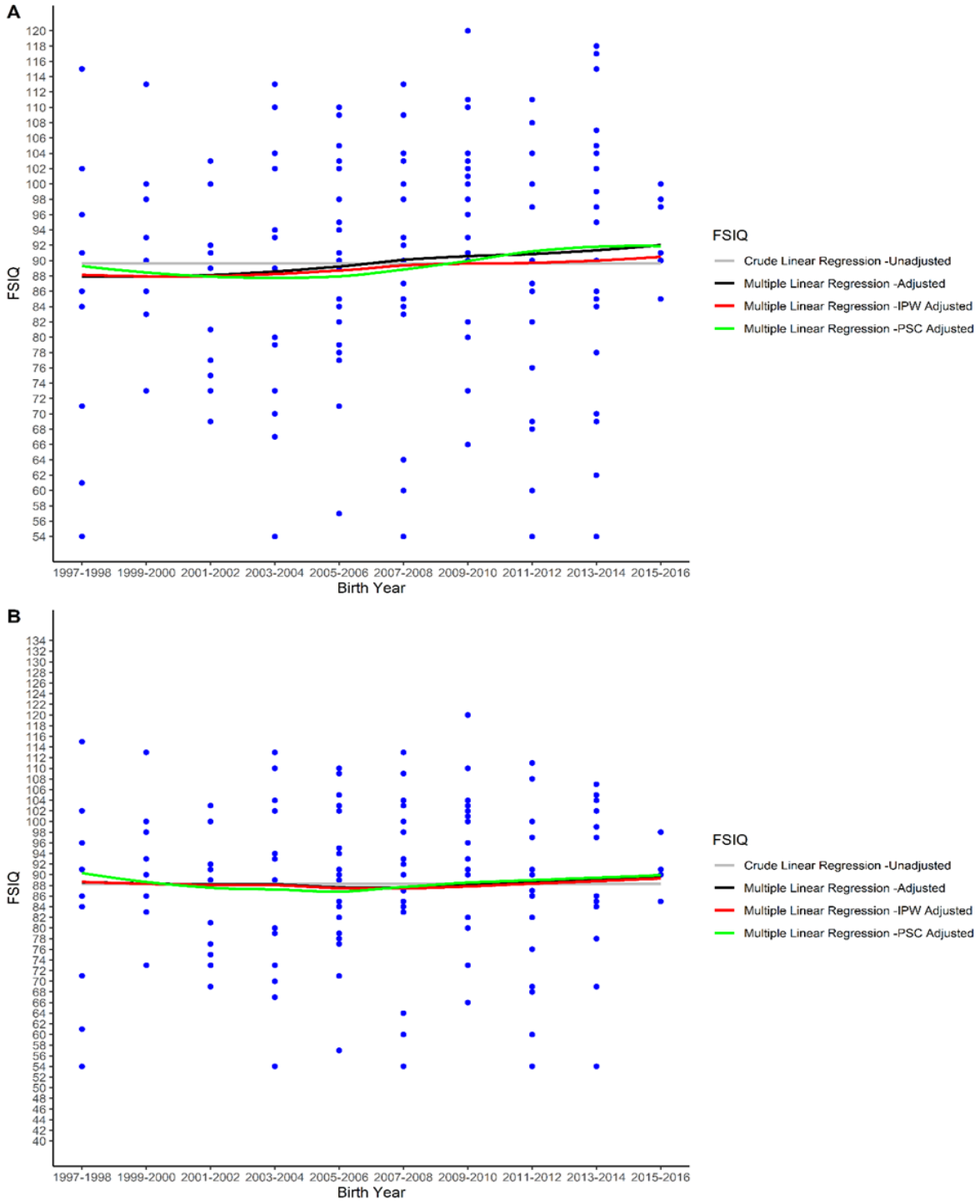


Figure 7 displays the VMI score trend:

1. VMI score trend for the entire sample.

Plot an on the whole sample captures a steady increase in IPW-adjusted and PSC-adjusted models. The time trend for multiple linear regression adjusted with P-value=0.02 and IPW-adjusted with p-value=0.037 was significant. Additionally, the P-value=0.147 for the time trend over the entire sample for the PSC-adjusted model.

2. VMI score trend in the severe group.

Plot B for the PSC-adjusted model frames a negligible decrease from 1997–1998 to 2001–2002, followed by a stable trend till 2005–2006 and experiencing a period of an upward trend, representing a P-value=0.398 for the time trend.

The IPW-adjusted model reveals a constant increase in the VMI score with the corresponding P-value=0.093 for the trend time.

The multiple linear regression adjusted model for the VMI score in the severe group showed a time trend with a P-value=0.073.

Figure 7: Linear trends of VMI score over birth years.

Plot A for the entire sample and plot B for the severe group.

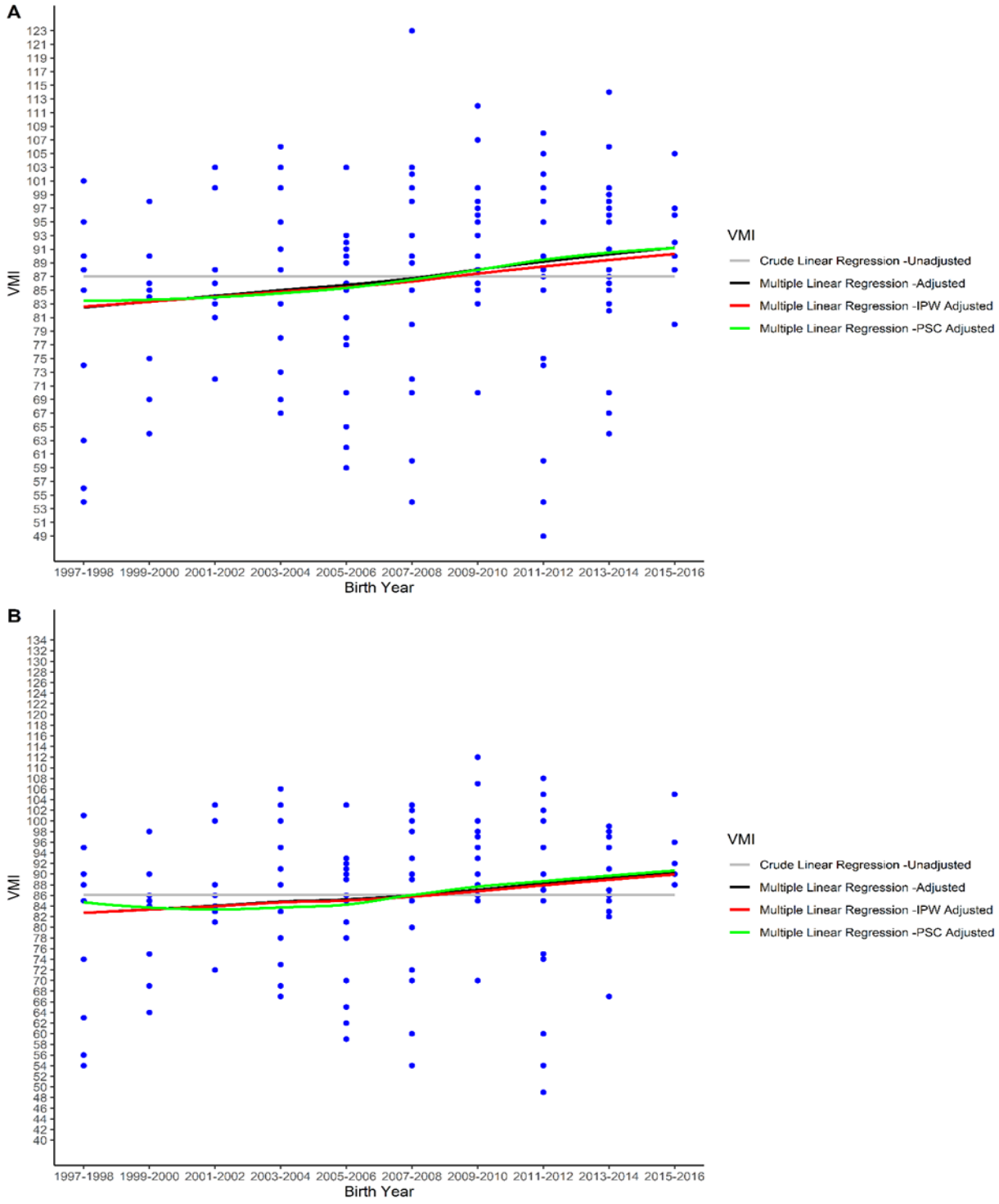


Figure 8 shows the ABAS GAC trend:

1. ABAS GAC score trend for the whole of the sample

Plot A on the whole sample depicts the gradual decrease in the trend of the score for both the PSC-adjusted and IPW-adjusted models; their corresponding P-values are 0.086 and 0.178, respectively.

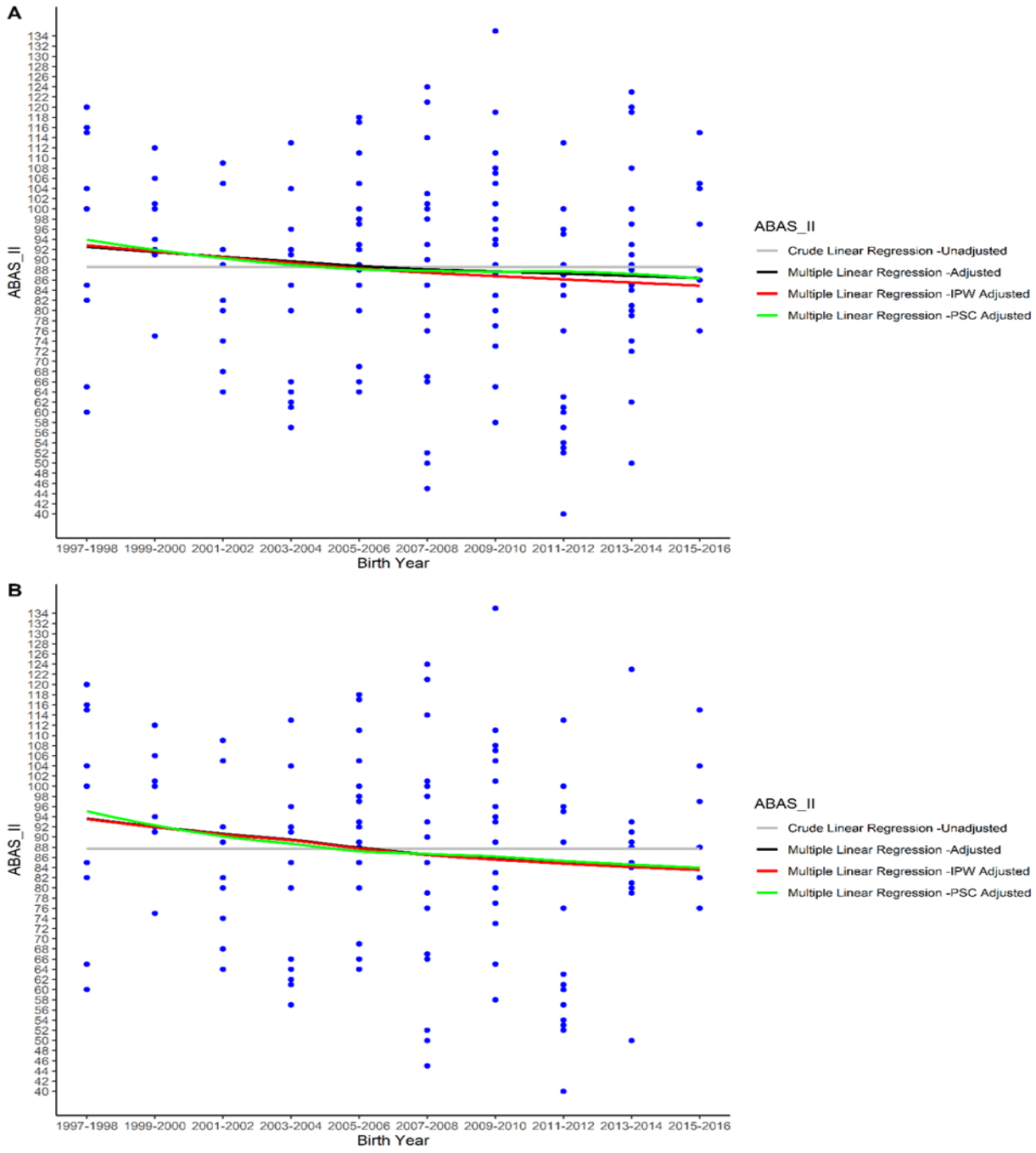
Furthermore, the multiple linear regression adjusted model for the ABAS GAC score in the entire group showed a time trend with the P-value=0.3.

2. ABAS GAC score trend in the severe group

Plot B portrays the constant decrease in the trend of the ABAS GAC score after adjusting for PS in both IPW-adjusted and PSC-adjusted models. However, the PSC-adjusted model had a significant time trend, with the corresponding P-value=0.042. The P-value=0.098 for the IPW-adjusted model was reported for the time trend, while the multiple linear regression adjusted model had a P-value=0.099.

Figure 8: Linear trends of ABAS-II score over birth years.

Plot A for the entire sample and plot B for the severe group.



Chapter 5: Discussion

5.1. Overview of our study

The heart pumps blood to the brain to grow and develop; this flow is enriched with nutrients and oxygen; Hence, inefficient blood circulation and low oxygen concentration may hinder brain development. Given this evidence-based fact, it should be no surprise that any cardiac issue may be interwoven with brain malfunction.

Accordingly, heart problems may pose problems for brain development, and consequently, newborns may mature neurologically later than healthy babies, which, in turn, might lead to delayed neurodevelopment.

In the last four to five decades, the prognosis for infants born with CHD has significantly improved, moving from a condition that, in most cases, resulted in death without diagnosis or treatment to one with an expected high survival rate (106,107).

Concerning these outstanding survival rates, it was anticipated that the neurocognitive trends would be similarly affected and there would be an improvement of delay in neurodevelopment.

Therefore, it created the false impression that brain function could be corrected concurrently with heart healing, reducing the prevalence of

adverse neurocognitive outcomes. The lack of changed neurocognitive outcomes dispelled this misunderstanding, though.

Hence, the increasing number of babies who survive CHD doesn't necessarily ensure improved neurological development.

It is possible that brain function, particularly in the neurocognitive aspect, will be indirectly affected according to the severity of the heart problem and illness due to survivorship bias.

Survivorship bias is a fallacy that clouds our perception of the true association between the effects of cardiac survival advances and neurocognitive outcomes. It arises when we need to consider prior failures and presume past success. It is critical to understand the whole narrative to dismiss this wrong estimation.

An initiative must be taken to lessen its impact on the relationship between the intervention and the outcome (108).

The gold standard for evaluating the influence of therapies, methods, and interventions in clinical trials is the RCT. However, RCTs are not always tractable because of various barriers, such as ethical concerns, financial constraints, and requirements for sample size and time (109).

To achieve balanced groups, reduce the number of variables in the model, mimic some features of RCT studies, address survivorship bias, and assess

the effectiveness of the intervention or treatment on the outcome while controlling for observed confounding variables, it may be helpful to summarise all confounding variables into one covariate, based on the PS analysis (110,111).

With the advent of the PS method, many confounders and selection biases can be tackled statistically.

5.2. Summary of our findings

This work investigates IPW and PSC techniques to diminish survivorship bias when evaluating neurocognitive and functional outcomes among CHD patients after cardiac surgery.

ICU LOS and death within 30 days were the surrogates for the PS calculation result. This time interval was identified in several studies as a substantial risk factor for adverse neurocognitive outcomes among children following heart surgery (112).

We also divided the infants into groups with a high and low degree of sickness, applying the KMC approach to PS. We discovered that deploying clustering analysis to PS can show more significant deviations of the trend from a flat line and provide a clearer picture of the neurocognitive and functional outcomes time trends.

Comparing trend lines on the entire sample with those in the severe group indicated that discriminating infants based on PS and clustering made trend lines clear to judge the overall trend.

Our results showed that between 1997 and 2016, for newborns distributed as having severe conditions, FSIQ scores overall showed a plateau trend, VMI scores steadily climbed, and ABAS GAC scores declined.

Although the overall trend in both the PSC-adjusted and IPW-adjusted models is the same, the PSC-adjusted model discloses more in-depth information about the volatility of the trend over time than the IPW-adjusted model. This model's underlying reason is smaller AIC and BIC than the IPW-adjusted model.

5.3. Reviewing other studies

5.3.1. Supporting studies

Regarding survivorship bias in a cohort study, the debate will arise about the existence of the healthy worker survivor effect (HWSE).

By definition, HWSE describes a biased selection process in which those who remain in the study tend to be healthier and face the most exposure than those who leave the study. In other words, those who leave the study sooner are sicker (113).

The HWSE can be conceptualized as bias due to time-varying confounding or selection bias leading to differences between baseline characteristics of the exposed and unexposed groups.

It is noted that changing HWSE over time may threaten an inference's validity (114).

Multiple methods have been implemented to overcome the HWSE and ensure the study's reliability.

Sometimes, it is advised to consider the first half of the cohort study or an early period to deal with the time-varying confounder because the study sensitivity may worsen or change in later periods. This would be efficient in the case of a large sample size.

However, there are statistical approaches to make sure that the time-dependent confounder or selection bias is appropriately addressed.

One of the highly recommended initiatives noted in lots of literature was an IPW approach that reweights observed data. It uses weights inversely proportional to the likelihood that each subject received the exposure, creating a pseudo-population in which time-dependent confounders no longer predict exposure (115,116).

In our study, the complexity and severity of the illness of CHD as the focal predictor had a time-varying nature. Thereby, the results might have been threatened by the presence of HWSE.

As we had to assign a PS to each person, using only an early cohort was not appropriate for our study because the study population was small, and just considering the first part of the cohort could result in losing some participants, so it was not feasible. Additionally, our team affirmed the list of predictors for developing the PS, which are pre-operative, intraoperative, and early post-operative; therefore, it addresses the issue of using early measurements to calculate the PS.

To take advantage of statistical methods for dealing with HWSE, the IPW was chosen because, on the one hand, it is widely used for calculating PS, and on the other hand, it can mask the effect of time-varying confounders; therefore, the influence of HWSE on changing the study reliability was potentially controlled.

The fact that we can only debug phenomena that arise after birth, like confounding factors, but not those that consolidate within the pregnancy, can have a sizable impact on brain development.

Multiple fetal imaging explorations have demonstrated that infants with CHD confront brain growth anomalies during gestation, primarily in the third trimester (117,118).

SV fetuses face lower maturation of grey matter as well (119).

Indisputably, the placenta plays a crucial role in fetal brain development. Otherwise, an impaired placenta can easily mask the oxygen flow, resulting in lower cerebral oxygen levels than other organs.

As a result, a normal placenta eases and accelerates cerebral blood flow and fosters proper brain development (120).

Despite the placenta's pivotal role in fetal brain evolution, a range of prenatal, surgical, and post-surgical components may disturb normal brain development in these target groups, entailing prolonged postnatal hypoxia before surgery (121,122), uneven anesthetic exposure (123), the time of CPB during surgery, postoperative complications during recovery, and lengthy hospitalization (124).

These aspects can negatively impact brain development and cause adverse long-term cognitive and functional outcomes.

Alton GY et al. conducted a study in 2010 that focused on the adaptive skill domains of six-month-old infants or younger who underwent Norwood and vascular procedures. They noted that the survivors of these palliative care efforts had delays in ABAS GAC scores over four times higher than the predicted rate for the general population. This study suggests that these

surgeries may be associated with a greater risk of experiencing developmental impediments measured using ABAS GAC (125).

In 2020, Atallah J. et al. undertook research that warned that children eligible for challenging three-stage surgical palliation for hypoplastic left heart syndrome after staying alive had VMI and ABAS GAC scores within 1 SD of the population mean. Their exploration also confirmed that 13% of the children had FSIQ scores below 70, and 25% had ABAS GAC scores below 70, designating a high-risk group (126).

Ricci M.F. et al. pursued an examination in 2018, which discussed that after the Norwood procedure for SV CHD, babies attained a decline in their ABAS II scores over time. Their work put forward the finding that children who had survived the surgery and were assessed when they were 4.5 years old or older showed a delay of 25% in their ABAS II scores, representing a deterioration in their functional abilities. This finding contradicts the widely held belief that children undergoing the Norwood procedure improve their practical skills as they age (127).

In 2015, Robertson C.M.T. et al. argued that the level of education attained by a child's mother might indicate their social context, and coping with this issue could lead to a better trajectory of neurocognitive skills. Moreover, the study highlighted preoperative lactate levels as a potentially modifiable

variable and stressed the importance of early diagnosis and aggressive resuscitation in young infants with CHD (128).

A body of research by Gaynor J.W. in 2014 examined the cognitive abilities of SV children at the age of six months. The investigation proposed that infants who had experienced more prolonged CPB and Norwood surgeries had lower FSIQ scores than those who had not. This result implied that the procedures' duration must be restricted; otherwise, it can even affect infants' cognitive outcomes (129).

In terms of deploying PS and KMC, there are countless kinds of research; here are some of them that support our hypothesis of using these methods to minimize survivorship bias and other confounding in observational studies:

It is common practice to estimate the causal effects of treatment on outcomes using observational studies. However, the distribution of baseline traits between participants who received treatment and those who did not frequently show systematic differences. As a result, outcomes between treatment groups cannot be directly compared.

Causal inference based on PS techniques was introduced to estimate causal effects in observational studies without the crucial role of RCT (130).

When comparing interventions across treatment areas, this technique reduces bias by accounting for the conditional probability of the treatment (131).

In logistic regression, the sample size is conceptualized as events per variable (EPV). The number of variables that should be included in a prediction model is not predetermined and frequently depends on several factors. The 'one in ten rules' limits the number of variables or parameters estimated from data, a common practice in traditional clinical prediction modeling. This rule states that a model can consider one variable for every ten events (132).

1996 Peduzzi P. et al. suggested multiple regression may yield inaccurate coefficients or effect sizes when the EPV numbers are less than 10. The one in ten rule states that if the EPV values decrease, it is more likely to impose biased coefficients in regression models. The analysis will be threatened by overfitting. Hence, combining multiple variables into one variable through PS can push the boundaries. It is feasible when the sample size is insufficient for each covariate to have at least ten samples (133).

Elze, M.C. et al., in a study published in 2017, compared the four approaches to PS. The results revealed that in most PS approaches, data from all study participants were retained, providing excellent covariate balance in most circumstances (134).

Cepeda M.S. (2003), by assigning a new cut-off point for EPV, advocated the PS method and stated that analysis data based on the PS technique provides a more unbiased and precise perspective with more empirical power than conventional logistic regression when the number of EPV is not enough to meet the cut point of seven because, for achieving accurate and reliable results by setting up logistic regression, we need at least eight events per confounder, which is not possible all the time (135).

A practical extension to PS was proposed by Imbens (2000), named generalized propensity score (GPS), to expand the PS methodology for binary treatments to multivalued and continuous treatments since all the treatment variables mentioned beforehand were binary (136).

By taking advantage of the GPS concept, Chunhao T (2013) discovered differences in the performance of four clustering methods. KMC, model-based clustering, Fuzzy C-means clustering, and partitioning around medoids were put into practice. The conclusion suggested that the KMC algorithm outperformed the other three methods. Therefore, applying the KMC algorithm to similar groups of subjects based on their transformed GPS is recommended (137).

D'Attoma et al. (2019) pointed out that both PS and cluster-based methods can effectively achieve unbiased results in quasi-experimental studies and undermine the effect of selection bias in non-randomized investigations.

Combining these methods provides a means of examining bias reduction, case retention, and covariate balance—the cluster-based technique narrows at least as much bias as the PS methods. While the PS method can deal with preferences better than the cluster-based method under certain conditions, the cluster-based approach is more advantageous under other circumstances (138).

Amoah et al. (2020) researched to compare PS and traditional regression analysis efficacy in estimating causal effects in the observational study. The results suggest PS analysis is more advantageous than traditional regression analysis. However, PS analysis cannot address the wrong study design or data inaccuracy (139).

5.3.2. Opposed studies.

However, some studies are against exploiting PS and express no benefit from PS adjustments. They are as follows:

Stürmer T. (2006) argued that although PS methods have become more dominant and widespread, there is limited evidence that these methods can produce distinct differences and unprejudiced forecasts compared to conventional multivariable methods. This study has not found any advantage in using PS adjustments over conventional regression methods in reducing bias or improving precision in nonrandomized studies (140).

Shah, B. R. (2005) claimed that the findings of observational studies were similar, regardless of whether conventional regression or PS were used to adjust for confounding. (141)

Even in research published in peer-reviewed surgical journals, several studies report their PS methodology insufficiently, which may result in a lack of accuracy and validity in the study's conclusion, according to a study by Grose E et al. (2020) that also revealed the presence of publication bias (142).

Cook et al. (1989) proposed that analyzing data by stratifying subjects based on a PS is less influenced by the strong correlation between exposure and confounding variables than analyzing data using a multivariate confounding score (143).

5.4. Strength and limitation

5.4.1. Strengths

In addition to the many merits mentioned before for applying PS in all the non-randomized control studies, this method has more advantages.

1. Missing data may be addressed with this method. Complete case analysis,

the missing indicator method, multiple imputations, and combining multiple

imputations and the missing indicator method are some approaches for handling missing values through PSM and IPW (144).

2. PS can produce accurate estimates of treatment effects in small study sample sizes or low treatment prevalence while including the true confounders (the variables related only to the outcome).

Besides, this application reduces the list of variables to a single score and only needs one degree of freedom for all the covariates in the model; it can be helpful in situations with small sample sizes and poor power to detect differences between groups (145).

3. Since PS attempts to mimic randomization, this novel method

Helps analyze rare outcomes and exposure. Traditional covariate adjustment fails to produce enough data in the case of rare outcomes to ensure the validity of the results (146).

5.4.2. Limitations

We built our hypothesis on two assumptions.

First, confounding variables have been measured, and there are no unmeasured confounders.

Second, each participant has a nonzero probability of receiving each treatment. Regarding the named assumption, the concern may arise around the following aspect of holding or violating it (147).

There might be a better way to form a PS analysis to account for survivorship bias if all the following questions are answered precisely.

1. How were the variables for running the regression selected for the PS model?

There needs to be more said regarding the issue of variable selection for PS models in the epidemiology literature. To our knowledge, no proper guidelines exist for selecting factors relevant to PS calculations.

Some studies suggested that it would be better to include variables unrelated to the exposure but linked to the outcome when selecting the variable for PS modelling. Including these factors will reduce the estimated exposure effect's variance and decrease bias (148).

2. How much could unmeasured or residual confounding have influenced the results?

From an epidemiologic point of view, residual confounding stems from the measurement error of a covariate, while unmeasured confounding refers to confounding that was mistakenly dropped from the model.

Lots of studies have highlighted that PS methods, such as the traditional multivariable regression method cannot mask the effect of unmeasured or incorrectly measured covariates (149).

The study results may need to be more balanced due to unobserved confounding variables that need to be investigated more (150).

Incorrect PS model specification could prohibit attaining an adequate balance, resulting in residual confounding bias. Employing proper diagnostics to evaluate the PS and make sure that it has adequately minimized. confounding bias is a crucial component of PS deployment (151,152).

Using diagnostics in the PS is the way to trust the result more. Applying the proper diagnostics may show the balance of potential confounders attained by the PS (153,154).

Based on this extension, there might be a better initiative on PS to control the survivorship bias.

3. How much did the treatment groups overlap?

While the conventional IPW we used and other PS strategies, such as matching can account for disparities in measured attributes, these techniques may have drawbacks regarding the target population, balance, and precision (155).

Areas of non-overlap in covariate distributions can be found by comparing PS distributions between treatment groups, which is frequently missed when using the PS (156).

Overlap weighting is a PS technique that tries to replicate key RCT characteristics. Similar patient features across treatments are referred to as balance, and this is a crucial requirement to prevent bias. Precision is the confidence interval (CI) in assessing the relationship between the intervention and the outcome; more accurate estimates have a narrower CI and higher statistical power. Therefore, adjusting based on conventional IPW might not precisely control for survivor bias.

4. How was balance assessed?

Lack of similarity in the covariate distributions between treatment groups frequently makes estimating the average treatment effects difficult. Because this may produce inaccurate estimates (157).

To ensure that the covariate balance across the treatment groups is accomplished, trimming the tails of the PS distribution is strongly advised to develop the IPW estimators and also reduce bias by unmeasured confounders (158,159).

Chapter 6: Conclusion

6.1. Summary

Based on the database made available for this study, which was a part of the Western Canadian CPTFP, after applying PS adjustment, the results revealed that in babies with severe CHD, over time, FSIQ scores overall remained the VMI scores steadily climbed, and ABAS GAC scores declined in some models.

Although it can be discussed that the FSIQ score followed a period of instability over time, its overall trend shows no changes. To ensure a genuine trend with an in-depth outlook. It can be seen that the changes in FSIQ scores over time in each discrete period are not more than ½ SD, so the ups and downs in trends cannot be considered clinically significant changes. The results remained plateaued, and we witnessed a lack of improvement or declined in the FSIQ score.

It would be good news if no declining trends in the FSIQ score were adjusted for the PS over time. It means that, despite sicker patients, there was no deterioration in the FSIQ score.

Furthermore, an improving trend for VMI scores over time was reported

after adjusting for the PS, which is good news, which means in the domain and skills covered by the VMI score, these babies had better performance.

Regarding the trends in FSIQ and VMI scores, it can be said that palliative surgical cardiac surgery helped them survive heart problems and positively influenced these neurocognitive domains.

However, there was a trend of decreasing ABAS GAC scores over time, which is concerning and requires future study to determine modifiable variables that may be particularly associated with the ABAS GAC score.

To find the significance of trend lines over time, we include the birth year or time in the linear regression model to ensure an independent association between the period and PS to find the significance of trend lines over time. In other words, these two are not collinear. In addition, by including the time variable in the linear regression models, the constant sensitivity of the analysis over time was statistically supported.

The PS adjustment methods with clustering have the advantage of more accurate estimations, more explicit trends showing more significant deviations from a flat slope, and, therefore, improved interpretations of outcome time trends.

Ultimately, this study's findings will provide new knowledge for future targeted treatments that seek to lessen the burden of illness before children enter school.

6.2. Future research and recommendations

We know from our prior work on disseminating the findings for the advancement of statistical methods that software codes alone are insufficient to effectively reach and assist many scientists, thereby putting this named approach into practice through divergent study populations like adults, with different sample sizes, totally different predictors, and contrasting biases are highly recommended.

Without a doubt, these modified techniques can be used in various biomedical studies where longitudinal outcomes are of interest, adding different extensions to them, like diagnostics, trimming, etc.

References

1. Meller CH, Grinenco S, Aiello H, Córdoba A, Sáenz-Tejeira MM, Marantz P, Otaño L. Congenital heart disease, prenatal diagnosis, and management. Arch Argent Pediatr. 2020 Apr;118(2):e149-e161. English, Spanish. doi: 10.5546/aap.2020.eng.e149. PMID: 32199055.

2. Health Canada. Congenital Anomalies in Canada — A Perinatal Health Report, 2002. Ottawa: Minister of Public Works and Government Services Canada; 2002.

3. Dolbec K, Mick NW. Congenital heart disease. Emerg Med Clin North Am. 2011 Nov;29(4):811-27, vii. doi: 10.1016/j.emc.2011.08.005. PMID: 22040709.

4. Sun R, Liu M, Lu L, Zheng Y, Zhang P. Congenital Heart Disease: Causes, Diagnosis, Symptoms, and Treatments. Cell Biochem Biophys. 2015 Jul;72(3):857-60. doi: 10.1007/s12013-015-0551-6. PMID: 25638345.

5.McCormick AD, Wilde MM, Charpie CE, Saylor KM, Yu S, Schumacher KR, Cousino MK. Psychological functioning in paediatric patients with single ventricle heart disease: a systematic review. Cardiol Young. 2022 Feb;32(2):173-184. doi: 10.1017/S1047951122000063. Epub 2022 Jan 20. PMID: 35045915.

6. Greaney D, Honjo O, O'Leary JD. The single ventricle pathway in pediatrics for anesthetists. BJA Educ. 2019 May;19(5):144-150. doi: 10.1016/j.bjae.2019.01.008. Epub 2019 Mar 6. PMID: 33456883; PMCID: PMC7807966.

7. Rao PS, Agarwal A. Advances in the Diagnosis and Management of Congenital Heart Disease in Children. Children (Basel). 2022 Jul 15;9(7):1056. doi: 10.3390/children9071056. PMID: 35884040; PMCID: PMC9319796.

8. Gross RE. Surgical management of the patent ductus arteriosus: with a summary of four surgically treated cases. Ann Surg 1939;110:321-56. 10.1097/00000658-193909000-00001

9. congenital heart disease. Heart and Stroke Foundation of Canada http://www.heartandstroke.ca/heart/conditions/congenital-heart-disease– accessed June 20, 2017.

10. McCrindle BW, Williams RV, Mitchell PD, Hsu DT, Paridon SM, Atz AM, Li JS, Newburger JW; Pediatric Heart Network Investigators. Relationship of patient and medical characteristics to health status in children and adolescents after the Fontan procedure. Circulation. 2006 Feb 28;113(8):1123-9. doi: 10.1161/CIRCULATIONAHA.105.576660. Epub 2006 Feb 20. PMID: 16490823.

11. Shillingford AJ, Glanzman MM, Ittenbach RF, Clancy RR, Gaynor JW, Wernovsky G. Inattention, hyperactivity, and school performance in a population of school-age children with complex congenital heart disease. Pediatrics. 2008 Apr;121(4): e759-67. doi: 10.1542/peds.2007-1066. PMID: 18381503.

12.Kamphuis M, Ottenkamp J, Vliegen HW, Vogels T, Zwinderman KH, Kamphuis RP, Verloove-Vanhorick SP. Health-related quality of life and health status in adult survivors with previously operated complex congenital heart disease. Heart. 2002 Apr;87(4):356-62. doi: 10.1136/heart.87.4.356. PMID: 11907011; PMCID: PMC1767074.

13. Hansen BM, Dinesen J, Hoff B, Greisen G. Intelligence in preterm children at four years of age as a predictor of school function: a longitudinal controlled study. Dev Med Child Neurol. 2002 Aug;44(8):517-21. doi: 10.1017/s0012162201002481. PMID: 12206616.

14. Forbess JM, Visconti KJ, Bellinger DC, Jonas RA. Neurodevelopmental outcomes in children after the Fontan operation. Circulation. 2001 Sep 18;104(12 Suppl 1): I127-32. doi: 10.1161/hc37t1.094816. PMID: 11568043.

15. Mellion K, Uzark K, Cassedy A, et al. Health-related quality of life outcomes in children and adolescents with congenital heart disease. J Pediatr. 2014;164:781–9. doi: 10.1016/j.jpeds.2013.11.066.

16.Moyses Szklo, F. Javier Nieto. Beyond the basic fourth edition.

17. Marino BS, Lipkin PH, Newburger JW, Peacock G, Gerdes M, Gaynor JW, Mussatto KA, Uzark K, Goldberg CS, Johnson WH Jr, Li J, Smith SE, Bellinger DC, Mahle WT; American Heart Association Congenital Heart Defects Committee, Council on Cardiovascular Disease in the Young, Council on Cardiovascular Nursing, and Stroke Council. Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. Circulation. 2012 Aug 28;126(9):1143-72. doi: 10.1161/CIR.0b013e318265ee8a. Epub 2012 Jul 30. PMID: 22851541.

18. Mitchell SC, Korones SB, Berende HW. Congenital heart disease in 56,109 births. Incidence and natural history. Circulation. 1971; 43:323–32.

19. Mandalenakis Z, Rosengren A, Skoglund K, Lappas G, Eriksson P, Dellborg M. Survivorship in Children and Young Adults with Congenital Heart Disease in Sweden. JAMA Intern Med. 2017 Feb 1;177(2):224-230. doi: 10.1001/jamainternmed.2016.7765. PMID: 27992621.

20. Moons P, Bovijn L, Budts W, Belmans A, Gewillig M. Temporal trends in survival to adulthood among patients born with congenital heart disease from 1970 to 1992 in Belgium. Circulation. 2010 Nov 30;122(22):2264-72. doi: 10.1161/CIRCULATIONAHA.110.946343. Epub 2010 Nov 22. PMID: 21098444.

21. Waldhausen JA, Boruchow I, Miller WW, Rashkind WJ. Transposition of the great arteries with the ventricular septal defect. Palliation by atrial septostomy and pulmonary artery banding. Circulation. 1969 May;39(5 Suppl 1): I215-21. doi: 10.1161/01.cir.39. 5s1.i-215. PMID: 5792976.

22. Crafoord C, Ejrup B, Gladnikoff H. Coarctation of the aorta. Thorax. 1947; 2:121–152.

23. McMullan MH, Wallace RB, Weidman WH, McGoon DC. Surgical treatment of complete atrioventricular canal. Surgery. 1972 Dec;72(6):905-12. PMID: 5087275.

24. MUSTARD WT, KEITH JD, TRUSLER GA, FOWLER R, KIDD L. THE SURGICAL MANAGEMENT OF TRANSPOSITION OF THE GREAT VESSELS. J Thorac Cardiovasc Surg. 1964 Dec; 48:953-8. PMID: 14239984.

25. SENNING A. Surgical correction of transposition of the great vessels. Surgery. 1959 Jun;45(6):966-80. PMID: 13659340.

26. Rastelli GC, Wallace RB, Ongley PA. Complete repair of transposition of the great arteries with pulmonary stenosis. A review and report of a case corrected by using a new surgical technique. Circulation. 1969 Jan;39(1):83-95. doi: 10.1161/01.cir.39.1.83. PMID: 5782810.

27. Jatene AD, Fontes VF, Paulista PP, Souza LC, Neger F, Galantier M, Sousa JE. Anatomic correction of transposition of the great vessels. J Thorac Cardiovasc Surg. 1976 Sep;72(3):364-70. PMID: 957754.

28. Fontan F, Baudet E. Surgical repair of tricuspid atresia. Thorax. 1971 May;26(3):240-8. doi: 10.1136/thx.26.3.240. PMID: 5089489; PMCID: PMC1019078.

29. Dittrich H, Buhrer C, Grimmer I, Dittrich S, Abdul-Khaliq H, Lange PE. Neurodevelopment at one year of age in infants with congenital heart disease. Heart. 2003; 89:436–41.

30. Ballweg JA, Wernovsky G, Gaynor JW. Neurodevelopmental outcomes following congenital heart surgery. Pediatr Cardiol. 2007; 28:126–3.

31. Rogers BT, Msall ME, Buck GM, Lyon NR, Norris MK, Roland JM, et al.. Neurodevelopmental outcome of infants with hypoplastic left heart syndrome. J Pediat. (1995) 126:496–8. 10.1016/S0022-3476(95)70478-7.

32. Calderon J, Newburger JW, Rollins CK. Neurodevelopmental and Mental Health Outcomes in Patients With Fontan Circulation: A State-of-the-Art Review. Front Pediatr. 2022 Mar 9;10:826349. doi: 10.3389/fped.2022.826349. PMID: 35356444; PMCID: PMC8959547.

33. Gaynor JW, Nord AS, Wernovsky G, Bernbaum J, Solot CB, Burnham N, Zackai E, Heagerty PJ, Clancy RR, Nicolson SC, Jarvik GP, Gerdes M.

Apolipoprotein E genotype modifies the risk of behavior problems after infant cardiac surgery. Pediatrics. 2009 Jul;124(1):241-50. doi: 10.1542/peds.2008-2281. PMID: 19564306; PMCID: PMC2840402.

34.Donofrio MT, Bremer YA, Schieken RM, Gennings C, Morton LD, Eidem BW, Cetta F, Falkensammer CB, Huhta JC, Kleinman CS. Autoregulation of cerebral blood flow in fetuses with congenital heart disease: the brain sparing effect. Pediatr Cardiol. 2003 Sep-Oct;24(5):436-43. doi: 10.1007/s00246-002-0404-0. PMID: 14627309.

35. Saleem SN. Fetal MRI: An approach to practice: A review. J Adv Res 2014; 5:507-23. 10.1016/j.jare.2013.06.001

36. De Keersmaecker B, Claus F, De Catte L. Imaging the fetal central nervous system. Facts View Vis Obgyn. 2011;3(3):135-49. PMID: 24753859; PMCID: PMC3991457.

37. Sun L, Macgowan CK, Sled JG, et al. Reduced fetal cerebral oxygen consumption is associated with smaller brain sizes in fetuses with congenital heart disease. Circulation. 2015; 131:1313–1323.

38. Malamateniou C, Malik SJ, Counsell SJ, et al. Motion-compensation techniques in neonatal and fetal MR imaging. Am J Neuroradiol 2013; 34:1124-36. 10.3174/ajnr. A3128

39. Griffiths PD, Mousa HA, Finney C, et al. An integrated into utero MR method for assessing structural brain abnormalities and measuring intracranial volumes in fetuses with congenital heart disease: results of a prospective case-control feasibility study. Neuroradiology 2019; 61:603-11. 10.1007/s00234-019-02184-2

011; 3:135-49.

40. Beca J, Gunn JK, Coleman L, Hope A, Reed PW, Hunt RW, Finucane K, Brizard C, Dance B and Shekerdemian LS (2013). "New white matter brain injury after infant heart surgery is associated with the diagnostic group and the use of circulatory arrest." Circulation 127: 971–979.

41. Dimitropoulos A, McQuillen PS, Sethi V, Moosa A, Chau V, Xu D, Brant R, Azakie A, Campbell A, Barkovich AJ, Poskitt KJ and Miller SP (2013). "Brain injury and development in newborns with critical congenital heart disease." Neurology 81(3): 241–248.

42. Newburger JW, Jonas RA, Wernovsky G, Wypij D, Hickey PR, Kuban KCK, Farrell DM, Holmes GL, Helmers SL, Constantinou J, Carrazana E, Barlow JK, Walsh AZ, Lucius KC, Share JC, Wessel DL, Hanley FL, Mayer JE, Castaneda AR, Ware JH. Comparing the perioperative neurologic effects of hypothermic circulatory arrest versus low-flow cardiopulmonary bypass in infant heart surgery. N Eng J Med. 1993; 329:1057–64.

43. Miller SP, McQuillen PS, Hamrick S, Xu D, Glidden DV, Charlton N, Karl T, Azakie A, Ferriero DM, Barkovich AJ, Vigneron DB. Abnormal brain development in newborns with congenital heart disease. N Engl J Med. 2007 Nov 8;357(19):1928-38. doi: 10.1056/NEJMoa067393. PMID: 17989385.

44. Mahle WT, Tavani F, Zimmerman RA, et al. An MRI study of neurological injury before and after congenital heart surgery. Circulation 2002 ;106(12 Suppl 1): I109–14.

45. Gaynor JW, Wernovsky G, Jarvik GP, Bernbaum J, Gerdes M, Zackai E, Nord AS, Clancy RR, Nicolson SC, Spray TL. Patient characteristics are important determinants of neurodevelopmental outcomes one year after neonatal and infant cardiac surgery. J Thorac Cardiovasc Surg. 2007 May;133(5):1344-53, 1353.e1-3. doi: 10.1016/j.jtcvs.2006.10.087. Epub 2007 Mar 19. PMID: 17467455; PMCID: PMC2844117.

46. Naim MY, Gaynor JW, Chen J, et al. Subclinical seizures identified by postoperative electroencephalographic monitoring are common after neonatal cardiac surgery. J Thorac Cardiovasc Surg. 2015; 150:169–178.

47. Kornilov IA, Sinelnikov YS, Soinov IA, et al. Outcomes after aortic arch reconstruction for infants: deep hypothermic circulatory arrest versus moderate hypothermia with selective antegrade cerebral perfusion. Eur J Cardiothorac Surg. 2015;48: e45–50.

48. Hövels-Gürich HH, Seghaye MC, Däbritz S, Messmer BJ, von Bernuth G. Cognitive and motor development in preschool and school-aged children after neonatal arterial switch operation. J Thorac Cardiovasc Surg. 1997 Oct;114(4):578-85. doi: 10.1016/S0022-5223(97)70047-3. PMID: 9338643.

49. Wernovsky G, Licht DJ. Neurodevelopmental Outcomes in Children with Congenital Heart Disease-What Can We Impact? Pediatr Crit Care Med. 2016 Aug;17(8 Suppl 1): S232-42. doi: 10.1097/PCC.0000000000000800. PMID: 27490605; PMCID: PMC4975480.

50. Bellinger DC, Wypij D, du Plessis AJ, Rappaport LA, Jonas RA, Wernovsky G, Newburger JW. Neurodevelopmental status at eight years in children with dextro-transposition of the great arteries: the Boston Circulatory Arrest Trial. J Thorac Cardiovasc Surg. 2003 Nov;126(5):1385-96. doi: 10.1016/s0022-5223(03)00711-6. PMID: 14666010.

51. Cheung PY, Chui N, Joffe AR, Rebeyka IM, Robertson CM; Western Canadian Complex Pediatric Therapies Project, Follow-up Group. Postoperative lactate concentrations predict the outcome of infants aged 6 weeks or less after intracardiac surgery: a cohort follow-up to 18 months. J Thorac Cardiovasc Surg. 2005 Sep;130(3):837-43. doi: 10.1016/j.jtcvs.2005.04.029. PMID: 16153937.

52. Kaleelullah RA, Garugula N. Teratogenic Genesis in Fetal Malformations. Cureus. 2021 Feb 5;13(2):e13149. doi: 10.7759/cureus.13149. PMID: 33692919; PMCID: PMC7937350.

53. Bale TL. Epigenetic and transgenerational reprogramming of brain development. Nat Rev Neurosci. 2015 Jun;16(6):332-44. doi:

54. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Third Edition Lippincott-Williams-Wilkins Publishers; Philadelphia, PA: 2008.

55. Haneuse S. Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research. Med Care. 2016 Apr;54(4):e23-9. doi: 10.1097/MLR.0000000000000011. PMID: 24309675; PMCID: PMC4043938.

56. Glesby MJ, Hoover DR. Survivor treatment selection bias in observational studies: examples from the AIDS literature. Ann Intern Med. 1996 Jun 1;124(11):999-1005. doi: 10.7326/0003-4819-124-11-199606010-00008. PMID: 8624068.

57. Starks H, Diehr P, Curtis JR. The challenge of selection bias and confounding in palliative care research. J Palliat Med. 2009 Feb;12(2):181-7. doi: 10.1089/jpm.2009.9672. PMID: 19207063; PMCID: PMC2982715.

58. Yang JY, Webster-Clark M, Lund JL, Sandler RS, Dellon ES, Stürmer T. Propensity score methods to control for confounding in observational cohort studies: a statistical primer and application to endoscopy research.

Gastrointest Endosc. 2019 Sep;90(3):360-369. doi: 10.1016/j.gie.2019.04.236. Epub 2019 Apr 30. PMID: 31051156; PMCID: PMC6715456.

59. Webster-Clark M, Stürmer T, Wang T, Man K, Marinac-Dabic D, Rothman KJ, Ellis AR, Gokhale M, Lunt M, Girman C, Glynn RJ. Using propensity scores to estimate effects of treatment initiation decisions: State of the science. Stat Med. 2021 Mar 30;40(7):1718-1735. doi: 10.1002/sim.8866. Epub 2020 Dec 29. PMID: 33377193.

60. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomized trials of statins and primary prevention of coronary heart disease. Stat Methods Med Res. 2013 Feb;22(1):70-96. doi: 10.1177/0962280211403603. Epub 2011 Oct 19. PMID: 22016461; PMCID: PMC3613145.

61. Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, Fremes SE. A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. Can J Cardiol. 2016 Feb;32(2):259-65. doi: 10.1016/j.cjca.2015.05.015. Epub 2015 May 23. PMID: 26315351.

62. D'Agostino RB Jr. Propensity score methods for bias reduction in comparing a treatment to a non-randomized control group. Stat Med. 1998; 17: 2265–228

63. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: From naïve enthusiasm to intuitive understanding. Statistical Methods in Medical Research. 2011;21(3):273-93.

64. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg. 2018 Jun 1;53(6):1112-1117. doi: 10.1093/ejcts/ezy167. PMID: 29684154.

65. Kane LT, Fang T, Galetta MS, Goyal DKC, Nicholson KJ, Kepler CK, Vaccaro AR, Schroeder GD. Propensity Score Matching: A Statistical Method. Clin Spine Surg. 2020 Apr;33(3):120-122. doi: 10.1097/BSD.0000000000000932. PMID: 31913173.

66. Cochran W.G. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 1968;24:295–313.

67. Huppler Hullsiek K., Louis T.A. Propensity score modeling strategies for the causal analysis of observational data. Biostatistics. 2002;3:179–193.

68. Raad H, Cornelius V, Chan S, Williamson E, Cro S. An evaluation of inverse probability weighting using the propensity score for baseline covariate adjustment in smaller population randomized controlled trials with a continuous outcome. BMC Med Res Methodol. 2020 Mar 23;20(1):70. doi: 10.1186/s12874-020-00947-7. PMID: 32293286; PMCID: PMC7092449.

69. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3):399-424. doi: 10.1080/00273171.2011.568786. Epub 2011 Jun 8. PMID: 21818162; PMCID: PMC3144483.

70. Jung YG, Kang MS, Heo J. Clustering performance comparison using K-means and expectation maximization algorithms. Biotechnol Biotechnol Equip. 2014 Nov 14;28(sup1): S44-S48. doi: 10.1080/13102818.2014.949045. Epub 2014 Nov 6. PMID: 26019610; PMCID: PMC4433949.

71. Leyrat C, Caille A, Donner A, Giraudeau B. Propensity score methods for estimating relative risks in cluster randomized trials with low-incidence binary outcomes and selection bias. Stat Med. 2014 Sep 10;33(20):3556-75. doi: 10.1002/sim.6185. Epub 2014 Apr 27. PMID: 24771662.

72. Creighton DE, Robertson CM, Sauve RS, Moddemann DM, Alton GY, Nettel-Aguirre A, Ross DB, Rebeyka IM; Western Canadian Complex Pediatric Therapies Follow-up Group. Neurocognitive, functional, and health outcomes at five years of age for children after complex cardiac surgery at six weeks or younger. Pediatrics. 2007 Sep;120(3):e478-86. doi: 10.1542/peds.2006-3250. PMID: 17766491.

73. Hurks P, Hendriksen J, Dek J, Kooij A. Accuracy of Short Forms of the Dutch Wechsler Preschool and Primary Scale of Intelligence: Third Edition. Assessment. 2016 Apr;23(2):240-9. doi: 10.1177/1073191115577189. Epub 2015 Mar 24. PMID: 25804438.

74.Sherwell S, Reid SM, Reddihough DS, Wrennall J, Ong B, Stargatt R. Measuring intellectual ability in children with cerebral palsy: can we do better? Res Dev Disabil. 2014 Oct;35(10):2558-67. doi: 10.1016/j.ridd.2014.06.019. Epub 2014 Jul 5. PMID: 25005064.

75. Sutton GP, Barchard KA, Bello DT, Thaler NS, Ringdahl E, Mayfield J, Allen DN. Beery-Buktenica Developmental Test of Visual-Motor Integration Performance in Children with traumatic brain injury and attention-deficit/hyperactivity disorder. Psychol Assess. 2011 Sep;23(3):805-9. doi: 10.1037/a0023370. PMID: 21875221.

76. Beery KE. Beery VMI: The Beery-Buktenica developmental test of visual-motor integration. Minneapolis, MN: Pearson. 2004.

77. Oakland T, Algina J. Adaptive behavior assessment system-II parent/primary caregiver form: Ages 0–5: Its factor structure and other implications for practice. Journal of Applied School Psychology. 2011;27(2):103-17.

78. Dupuis A, Moon MJ, Brian J, Georgiades S, Levy T, Anagnostou E, Nicolson R, Schachar R, Crosbie J. Concurrent Validity of the ABAS-II Questionnaire with the Vineland II Interview for Adaptive Behavior in a Pediatric ASD Sample: High Correspondence Despite Systematically Lower Scores. J Autism Dev Disord. 2021 May;51(5):1417-1427. doi: 10.1007/s10803-020-04597-y. PMID: 32776267.

79. Harrison PL, Oakland T. Adaptive Behavior Assessment System. 2. San Antonio, TX: Harcourt Assessment, Inc; 2003.

80. Harrison PL, Oakland T. ABAS-II assessment methods. Adaptive behavior assessment system-II: Elsevier; 2008. p. 37-49.

81. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003 May;41(5):582-92. doi: 10.1.097/01.MLR.0000062554.74615.4C. PMID: 12719681.

82. Karim H, Tara SM, Etminani K. Factors associated with length of hospital stay: A systematic review. J Health Biomed Inform. 2015;1:131–42.

83. Setareh M, Alavi NM, Atoof F. Severity of illness affecting the length of stay and outcomes in patients admitted to intensive care units, Iran, 2019. J Educ Health Promot. 2021 May 20;10:142. doi: 10.4103/jehp.jehp_780_20. PMID: 34222517; PMCID: PMC8224510.

84. Balasubramanian V, Suri JC, Ish P, Gupta N, Behera D, Gupta P, Chakrabarti S. Neurocognitive and Quality-of-life Outcomes Following Intensive Care Admission: A Prospective 6-month Follow-up Study. Indian J Crit Care Med. 2020 Oct;24(10):932-937. doi 10.5005/jp-journals-10071-23576. PMID: 33281317; PMCID: PMC7689110.

85. Madurski C, Treble-Barna A, Fink EL. Cognitive Impairment Following Pediatric Critical Illness: Time to Pay Attention. Pediatr Crit Care Med. 2018 Mar;19(3):277-278. doi: 10.1097/PCC.0000000000001436. PMID: 29499029; PMCID: PMC5836494.

86. Quach J, Barnett T (2015) Impact of chronic illness timing and persistence at school entry on child and parent outcomes: Australian longitudinal study. Acad Pediatr 15(1):89–95

87. Chiorean A, Savoy C, Beattie K, el Helou S, Silmi M, Van Lieshout R (2020) Childhood and adolescent mental health of NICU graduates: an observational study. Arch Dis Child 105:684–689.

88. Chitra P, Jeenu K (2016) Perspectives of strange environmental stressors among hospitalized children at AIMS, Kochi. Indian Journal of Public Health Research & Development 7(1):266.

89. Kyösti E, Ala-Kokko T, Ohtonen P et al (2019) Strengths and difficulties questionnaire assessment of long-term psychological outcome in children after intensive care admission*. Pediatr Crit Care Med 20(11):e496–e502.

90. Tabib A, Abrishami SE, Mahdavi M, Mortezaeian H, Totonchi Z. Predictors of Prolonged Mechanical Ventilation in Pediatric Patients After Cardiac Surgery for Congenital Heart Disease. Res Cardiovasc Med. 2016 Jul 20;5(3):e30391. doi: 10.5812/cardiovascmed.30391. PMID: 28105408; PMCID: PMC5219893.

91. Joung RH, Merkow RP. Is it Time to Abandon 30-Day Mortality as a Quality Measure? Ann Surg Oncol. 2021 Mar;28(3):1263-1264. doi 10.1245/s10434-020-09262-3. Epub 2021 Jan 3. PMID: 33393040; PMCID: PMC8148608.

92. Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. J Clin Epidemiol. 1999 Sep;52(9):885-92. doi: 10.1016/s0895-4356(99)00075-x. PMID: 10529029.

93. Kramer AA, Zimmerman JE. The relationship between hospital and intensive care unit length of stay. Crit Care Med. 2011 May;39(5):1015-22. doi: 10.1097/CCM.0b013e31820eabab. PMID: 21336128.

94. Ruttimann UE, Patel KM, Pollack MM. Length of stay and efficiency in pediatric intensive care units. Journal of Pediatrics. 1998 Jul;133(1):79–85.

95. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis: John Wiley & Sons; 2009.

96. Rencher AC, Schimek M. Methods of multivariate analysis. Computational Statistics. 1997;12(4):422-.

97. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa LdF, et al. Clustering algorithms: A comparative approach. PloS one. 2019;14(1): e0210236.

98. Sirin SR. Socioeconomic status and academic achievement: A meta-analytic review of research. Review of Educational Research. 2005; 75:417–453. doi: 10.3102/00346543075003417

99. Ko G, Shah P, Lee SK, Asztalos E. Impact of maternal education on cognitive and language scores at 18 to 24 months among extremely preterm neonates. Am J Perinatol. 2013 Oct;30(9):723-30. doi: 10.1055/s-0032-1331034. Epub 2012 Dec 27. PMID: 23271383.

100. Zhou Y, Müller HG, Zhu C, Chen Y, Wang JL, O'Muircheartaigh J, Bruchhage M, Deoni S; RESONANCE Consortium. Network evolution of regional brain volumes in young children reflects neurocognitive scores and

the mother's education. Sci Rep. 2023 Feb 20;13(1):2984. doi: 10.1038/s41598-023-29797-1. PMID: 36804963; PMCID: PMC9941570.

101. Jackson M, Kiernan K, McLanahan S. Maternal Education, Changing Family Circumstances, and Children's Skill Development in the United States and UK. Ann Am Acad Pol Soc Sci. 2017 Nov;674(1):59-84. doi: 10.1177/0002716217729471. Epub 2017 Oct 25. PMID: 29563643; PMCID: PMC5857959.

102. Newburger JW, Wypij D, Bellinger DC, du Plessis AJ, Kuban KC, Rappaport LA, Almirall D, Wessel DL, Jonas RA, Wernovsky G. Length of stay after infant heart surgery is related to the cognitive outcome at age eight years. J Pediatr. 2003 Jul;143(1):67-73. doi: 10.1016/S0022-3476(03)00183-5. PMID: 12915826.

103. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. European Heart Journal. 2011;32(14):1704-8.

104. Ciminata G, Geue C, Wu O, Deidda M, Kreif N, Langhorne P. Propensity score methods for comparative-effectiveness analysis: A case study of direct oral anticoagulants in the atrial fibrillation population. PLoS One. 2022 Jan 24;17(1): e0262293. doi: 10.1371/journal.pone.0262293. PMID: 35073380; PMCID: PMC8786176.

105. Kuha J. AIC and BIC: Comparisons of Assumptions and Performance. Sociological Methods & Research. 2004;33(2):188-229

106.Dodge-Khatami A. Advances and research in congenital heart disease. Transl Pediatr. 2016 Jul;5(3):109-111. doi: 10.21037/tp.2016.05.01. PMID: 27709090; PMCID: PMC5035758.

107.Gross RE. Surgical management of the patent ductus arteriosus: with a summary of four surgically treated cases. Ann Surg 1939;110:321-56. 10.1097/00000658-193909000-00001

108. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. Jama. 2016;316(17):1818-9.

109. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine. 2000;342(25):1887-92.

110. Haukoos JS, Lewis RJ. The Propensity Score. JAMA. 2015;314(15):1637-8.

111. Lalani N, Jimenez RB, Yeap B. Understanding propensity score analyses. International Journal of Radiation Oncology, Biology, Physics. 2020;107(3):404-7.

112. Karl TR, Hall S, Ford G, Kelly EA, Brizard CPR, Mee RBB, et al. Arterial switch with full-flow cardiopulmonary bypass and limited circulatory arrest:

Neurodevelopmental outcome. The Journal of Thoracic and Cardiovascular Surgery. 2004;127(1):213-22.

113. Richardson D, Wing S, Steenland K, McKelvey W. Time-related aspects of the healthy worker survivor effect. Ann Epidemiol. 2004;14:633–9.

114. Arrighi HM, Hertz-Picciotto I. The Evolving Concept of the Healthy Worker Survivor Effect. Epidemiology. 1994;5:189–96.

115. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. Am J Epidemiol. 2008;168:656–64.

116. Robins JM. Marginal Structural Models. 1997 Proc Am Stat Assoc. 1998:1–10.

117. Zeng S, Zhou Q, Zhou J, Li M, Long C, Peng Q. Volume of intracranial structures on three-dimensional ultrasound in fetuses with congenital heart disease. Ultrasound in Obstetrics & Gynecology. 2015;46(2):174-81.

118. Clouchoux C, Kudelski D, Gholipour A, Warfield SK, Viseur S, Bouyssi-Kobar M, et al. Quantitative in vivo MRI measurement of cortical development in the fetus. Brain structure & function. 2012;217(1):127-39.

119. Rajagopalan V, Votava-Smith JK, Zhuang X, Brian J, Marshall L, Panigrahy A, et al. Fetuses with single ventricle congenital heart disease manifest impairment of regional brain growth. Prenat Diagn. 2018;38(13):1042-8.

120. Leon RL, Mir IN, Herrera CL, Sharma K, Spong CY, Twickler DM, et al. Neuroplacentology in congenital heart disease: placental connections to neurodevelopmental outcomes. Pediatric Research. 2021:1-8.

121. Lynch JM, Gaynor JW, Licht DJ, editors. Brain injury during the transition in newborns with congenital heart disease: hazards of the preoperative period. Seminars in pediatric neurology; 2018: Elsevier.

122. Mebius MJ, van der Laan ME, Verhagen EA, Roofthooft MT, Bos AF, Kooi EM. Cerebral oxygen saturation during the first 72 h after birth in infants diagnosed prenatally with congenital heart disease. Early human development. 2016; 103:199-203.

123. Diaz LK, Gaynor JW, Koh SJ, Ittenbach RF, Gerdes M, Bernbaum JC, et al. Increasing cumulative exposure to volatile anesthetic agents is associated with poorer neurodevelopmental outcomes in children with hypoplastic left heart syndrome. The Journal of Thoracic and cardiovascular surgery. 2016;152(2):482-9.

124.Hoffman GM, Brosig CL, Bear LM, Tweddell JS, Mussatto KA. Effect of intercurrent operation and cerebral oxygenation on the developmental trajectory in congenital heart disease. The Annals of thoracic surgery. 2016;101(2):708-16.

125. Alton GY, Rempel GR, Robertson CM, Newburn-Cook CV, Norris CM. Functional outcomes after neonatal open cardiac surgery: comparison of survivors of the Norwood staged procedure and the arterial switch operation. Cardiology in the Young. 2010;20(6):668-75.

126. Atallah J, Garcia Guerra G, Joffe AR, Bond GY, Islam S, Ricci MF, et al. Survival, neurocognitive, and functional outcomes after completion of staged surgical palliation in a cohort of patients with hypoplastic left heart syndrome. Journal of the American Heart Association. 2020;9(4):e013632.

127. Ricci MF, Martin B-J, Joffe AR, Dinu IA, Alton GY, Guerra GG, et al. Deterioration of functional abilities in children surviving the Fontan operation. Cardiology in the Young. 2018;28(6):868-75.

128. Alton GY, Taghados S, Joffe AR, Robertson CM, Dinu I. Prediction of preschool functional abilities after early complex cardiac surgery. Cardiology in the Young. 2015;25(4):655-62.

129. Gaynor JW, Ittenbach RF, Gerdes M, Bernbaum J, Clancy RR, McDonald-McGinn DM, et al. Neurodevelopmental outcomes in preschool survivors of the Fontan procedure. The Journal of Thoracic and cardiovascular surgery. 2014;147(4):1276-83. E5.

130. Lanza ST, Moore JE, Butera NM. Drawing causal inferences using propensity scores: a practical guide for community psychologists. Am J

Community Psychol. 2013 Dec;52(3-4):380-92. doi: 10.1007/s10464-013-9604-4. PMID: 24185755; PMCID: PMC4098642.

131. PAUL R. ROSENBAUM, DONALD B. RUBIN, The central role of the propensity score in observational studies for causal effects, Biometrika, Volume 70, Issue 1, April 1983, Pages 41–55, https://doi.org/10.1093/biomet/70.1.41

132. Chowdhury MZI, Turin TC. Variable selection strategies and their importance in clinical prediction modeling. Fam Med Community Health. 2020 Feb 16;8(1):e000262. doi: 10.1136/fmch-2019-000262. PMID: 32148735; PMCID: PMC7032893.

133. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373-9. doi: 10.1016/s0895-4356(96)00236-3. PMID: 8970487.

134. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Nichols M, Stone GW, Pocock SJ. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. J Am Coll Cardiol. 2017 Jan 24;69(3):345-357. doi: 10.1016/j.jacc.2016.10.060. PMID: 28104076.

135. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low, and there are multiple confounders. American Journal of Epidemiology. 2003;158(3):280-7.

136. Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika. 2000;87(3):706-10.

137.Tu C, Jiao S, Koh WY. Comparison of clustering algorithms on generalized propensity score in observational studies: a simulation study. Journal of Statistical Computation and Simulation. 2013;83(12):2206-18

138. D'Attoma I, Camillo F, Clark M. A comparison of bias reduction methods: clustering versus Propensity score Subclassification and Weighting. The Journal of Experimental Education. 2019;87(1):33-54.

139. Amoah J, Stuart EA, Cosgrove SE, Harris AD, Han JH, Lautenbach E, Tamma PD. Comparing Propensity Score Methods Versus Traditional Regression Analysis for Evaluating Observational Data: A Case Study Evaluating the Treatment of Gram-Negative Bloodstream Infections. Clin Infect Dis. 2020 Dec 3;71(9):e497-e505. doi: 10.1093/cid/ciaa169. PMID: 32069360; PMCID: PMC7713675.

140. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use

and advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. Journal of Clinical Epidemiology. 2006;59(5): 437.e1-.e24.

141. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. Journal of Clinical Epidemiology. 2005;58(6):550-9.

142. Grose E, Wilson S, Barkun J, Bertens K, Martel G, Balaa F, Abou Khalil J. Use of Propensity Score Methodology in Contemporary High-Impact Surgical Literature. J Am Coll Surg. 2020 Jan;230(1):101-112.e2. doi: 10.1016/j.jamcollsurg.2019.10.003. Epub 2019 Oct 28. PMID: 31672675.

143. Cook EF, Goldman L. Performance of significance tests based on stratification by a multivariate confounder score or propensity score. Journal of clinical epidemiology. 1989;42(4):317-24.

144. Missing data Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. Psychological Methods, 21(3), 427–445.

145. Pirracchio R, Resche-Rigon M, Chevret S. Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. BMC Med Res Methodol. 2012 May 30;12:70. doi: 10.1186/1471-2288-12-70. PMID: 22646911; PMCID: PMC3511219.

146. Ross ME, Kreider AR, Huang YS, Matone M, Rubin DM, Localio AR. Propensity Score Methods for Analyzing Observational Data Like Randomized Experiments: Challenges and Solutions for Rare Outcomes and Exposures. Am J Epidemiol. 2015 Jun 15;181(12):989-95. doi: 10.1093/aje/kwu469. Epub 2015 May 20. PMID: 25995287.

147. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41-55.

148. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15;163(12):1149-56. doi: 10.1093/aje/kwj149. Epub 2006 Apr 19. PMID: 16624967; PMCID: PMC1513192.

149. Okoli GN, Sanders RD, Myles P. Demystifying propensity scores. Br J Anaesth. 2014 Jan;112(1):13-5. doi: 10.1093/bja/aet290. PMID: 24318697; PMCID: PMC3854550.

150.Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. J Intern Med. 2014 Jun;275(6):570-80. doi: 10.1111/joim.12197. Epub 2014 Feb 13. PMID: 24520806; PMCID: PMC4037382.

151. Yao XI, Wang X, Speicher PJ, et al. Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies. J Natl Cancer Inst. 2017;109:djw323.

152. West SG, Cham H, Thoemmes F, et al. Propensity scores as a basis for equating groups: basic principles and application in clinical treatment outcome research. J Consult Clin Psychol. 2014;82:906.

153. Austin PC. Assessing balance in measured baseline covariates using many-to-one matching on the propensity score. Pharmacoepidemiol Drug Saf. 2008;17(12):1218–25.

154. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009;28(25):3083–107.

155. Thomas LE, Li F, Pencina MJ. Overlap Weighting: A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial. JAMA. 2020 Jun 16;323(23):2417-2418. doi: 10.1001/jama.2020.7819. PMID: 32369102.

156. Garrido MM, Kelley AS, Paris J, et al. Methods for constructing and assessing propensity scores. Health Serv Res. 2014;49:1701–20.

157. Stürmer T, Rothman KJ, Avorn J, et al.. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails

of the propensity score distribution—a simulation study. Am J Epidemiol. 2010;172(7):843–854.

158. Walker AM, Patrick AR, Lauer MS, et al.. A tool for assessing the feasibility of comparative effectiveness research. Comp Eff Res. 2013;2013(3):11–20.

PS trimming can mitigate the bias by unmeasured confounders by focusing on a subset of subjects with better treatment equipoise.

159. Crump RK, Hotz VJ, Imbens GW, et al.. We are dealing with limited overlap in the estimation of average treatment effects. Biometrika. 2009;96(1):187–199.