This is to certify that the
thesis entitled

A Systematic Review of Neurodevelopmental and Cognitive Outcomes in Premature Children Born with Germinal Matrix and Intraventricular Hemorrhage (GM/IVH)

presented by

Jun-tsui Fan

has been accepted towards fulfillment
of the requirements for the

M.S. degree in Epidemiology

Major Professor's Signature

8/2/06

Date

A SYSTEMATIC REVIEW OF NEURODEVELOPMENTAL AND COGNITIVE OUTCOMES IN PREMATURE CHILDREN WITH GERMINAL MATRIX AND INTRAVENTRICULAR HEMORRHAGE

By
Jun-tsui Fan

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Epidemiology

2006

ABSTRACT

A SYSTEMATIC REVIEW OF NEURODEVELOPMENTAL AND COGNITIVE OUTCOMES IN PREMATURE CHILDREN WITH GERMINAL MATRIX AND INTRAVENTRICULAR HEMORRHAGE

By

Jun-tsui Fan

To determine whether isolated GM/IVH is associated with neurodevelopmental and cognitive abnormalities, I adopt a systematic review approach that combines qualitative and quantitative methods in this thesis. MEDLINE and ISI Web of Knowledge are the main sources for literature searches. A ranking of importance for the selected studies is provided after combining the evaluation with causal criteria and study quality criteria. The studies of van de Bor et al. (1988, 1993, and 2004) are regarded as important studies according to the ranking. Weighted OR shows that children with GM/IVH have a 2.5 times higher risk of abnormal neurodevelopment and abnormal general cognition than those without. The differences in weighted means of neurodevelopmental and cognitive tests range from 1.2 to 3.3. The results suggest that GM/IVH may slightly impair children's special cognitive outcomes, while its effects on children's abnormal global cognition is not significant. In addition, GM/IVH in preterm infants leads to an increased risk of adversely affected neurodevelopmental performance. The effect of GM/IVH on school performances of children or adolescents is unfavorable.

This thesis is dedicated with love to
Moi-kien Yi, Li-Tai-moi Yi, Ngiuk-Siuk Yi,
Yang-tse Fan, Jun-huei Fan,
Kai-hui Chang, and Yen-yi Ho.

ACKNOWLEDGEMENTS

This thesis originates from a lot of people's wisdom, enthusiasm, and aspiration. I wrote it down with a great hope of contributing the organized knowledge and interesting discovery to people who want to do further relevant explorations. I feel very glad for the accomplishment of this thesis and have an immense gratitude to people who have helped, supported, and took care of this work.

From the bottom of my heart, I thank Dr. Paneth, my advisor, for your enlightened teaching, generous assistance, and admirable patience. In the past two years, my abilities, confidence, and enthusiasm for researches were all elevated miraculously because of working with you. I appreciate Dr. Karna and Dr. Breslau, my committee, for your valuable suggestions and nice comments. Thank you so much.

Thank Mom for always being my strongest supporter either for my spirit or my daily life. Without you, I cannot achieve this challenging mission. Thank grandfather and grandmother for your encouragement and love. Thank Jun-huei, my twin sister, for always making me happy and cheering me up. Thank Yang-tse, my brother, for taking care of our family and allowing me not to worry about anything. Thank Yen-yi for answering all my research questions. Finally, I would like to thank Kai-hui for editing and formatting my thesis, helping me prepare the defense, taking care of my daily life, and giving me perpetual love and happiness. I am sincerely grateful for all that you have done. Thank you very much.

TABLE OF CONTENTS

LIST OF TABLES VI
Chapter One Introduction 1
I. Research Purpose 1
II. Background 1
III. Definition of GMH/IVH 2
IV. Classification of IVH in the Premature Infant 4
V. Diagnosis of IVH 7
Chapter Two Methods of this Research 10
I. Hypothesis 10
II. Literature Search 10
1. Sources for the Literature Search 10
2. Determining Key Concepts and Generating Search Terms 14
3. Strategies for Performing MEDLINE/ PubMed 15
III. Literature Selection 16
1. Inclusion and Exclusion Criteria 16
2. Selection with Titles, Abstracts, and Texts of Literature 17
IV. Literature Analysis 18
1. Determining the Type of Literature Review 18
2. Methods for Summarizing Evidence 21
Chapter Three Results 26
I. Studies Identified 26
II. Introduction of Birth Cohorts 26
III. Measures and Classification of IVH 28
IV. Follow-up and Assessment Rates in Studies 29
V. Study Quality Evaluation 31
VI. Ranking of Importance for Selected Studies 35
VII. Frequencies and Odds Ratios of Developmental Abnormalities 36
VIII. Averages and Standard Deviations of Developmental Tests 38
IX. Overall Evaluation and Interpretation 39
Chapter Four Discussion 42
I. Evaluation of Causal Criteria 42
II. Limitations 44
III. Strengths 46
IV. Conclusions 47
V. Further Discussion 47
TABLES 50
BIBLIOGRAPHY 90

LIST OF TABLES

Table 1-A (1 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies without a Full-Term Control Group 50
Table 1-A (2 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies without a Full-Term Control Group 51
Table 1-B (1 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies with a Full-Term Control Group 52
Table 1-B (2 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies with a Full-Term Control Group 53
Table 2-A (1 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies without a Full-Term Control Group 54
Table 2-A (2 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies without a Full-Term Control Group 55
Table 2-B (1 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies with a Full-Term Control Group 56
Table 2-B (2 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies with a Full-Term Control Group 57
Table 3-A (1 of 2): Study Quality Evaluation 58
Table 3-A (2 of 2): Study Quality Evaluation 59
Table 3-B (1 of 2): Study Quality Evaluation 60
Table 3-B (2 of 2): Study Quality Evaluation 61
Table 4 (1 of 2): Ranking of Importance for Selected Studies 62
Table 4 (2 of 2): Ranking of Importance for Selected Studies 63
Table 5-A (1 of 2): Frequencies and Odds Ratios of Abnormal Outcomes in Studies without a Full-Term Control Group 64
Table 5-A (2 of 2): Frequencies and Odds Ratios of Abnormal Outcomes in Studies without a Full-Term Control Group 66
Table 5-B (1 of 2): Frequencies and Odds Ratios of Special Cognitive Outcomes in a Study with a Full-Term Control Group 68
Table 5-B (2 of 2): Frequencies and Odds Ratios of Special Cognitive Outcomes in a Study with a Full-Term Control Group 69
Table 6-A (1 of 2): Means and Standard Deviations of Developmental Tests in Studies without a Full-Term Control Group 70
Table 6-A (2 of 2): Means and Standard Deviations of Developmental Tests in Studies without a Full-Term Control Group 72
Table 6-B (1 of 2): Means and Standard Deviations of Cognitive Tests in Studies with a Full-Term Control Group 74
Table 6-B (2 of 2): Means and Standard Deviations of Cognitive Tests in Studies with a

Full-Term Control Group 75
Table 7-A Weighted OR for Abnormal Neurodevelopment 76
Table 7-B Weighted OR for General Cognition 76
Table 7-C Weighted Means for Neurodevelopment: Motor Abilities 77
Table 7-D Weighted Means for General Cognition 77
Table 7-E Weighted Means for Special Cognition - Language 78
Table 7-F Weighted Means for Special Cognition– Reasoning & Memory 78
Table 8 (1 of 2): Causal Criteria for Selected Studies 79
Table 8 (2 of 2): Causal Criteria for Selected Studies 80
Appendix Table 1 (1 of 2): Measurements and Classification of IVH 81
Appendix Table 1 (2 of 2): Measurements and Classification of IVH 83
Appendix Table 2 (1 of 2): Measurements of Outcomes 85
Appendix Table 2 (2 of 2): Measurements of Outcomes 87

Chapter One
Introduction

I. Research Purpose

Many studies have discussed the relationship between severe brain damage (germinal matrix and/or intraventricular hemorrhage (GM/IVH) with parenchymal lesions/ ventricular enlargement (PL/VE)) and developmental abnormalities in premature children. However, the influence of GM/IVH alone on the outcomes of premature children has been insufficiently studied. The purpose of this research is to compare the outcomes of premature surviving children with and without GM/IVH in order to determine whether isolated GM/IVH is associated with neurodevelopmental and cognitive abnormalities.

II. Background

Intracranial hemorrhage (ICH) may affect newborns of all gestational ages and is often clinically silent. Germinal matrix hemorrhage and intraventricular hemorrhage (GMH/IVH) are most common in the premature population. Estimates of GMH/IVH frequency in preterm infants have declined from 30 to 40% in the early 1980s to less than 20% in the 1990s [1]. GM/IVH was once found in about 40% of very low birth weight (VLBW; birthweight< 1500 grams) infants [2]-[7]. However, a recent study suggests that 25-35% of extremely low birth weight (ELBW; birthweight< 1000 grams) infants will have some degrees of IVH identified in the early newborn period [8], and the prevalence in infants whose gestational age are less than 32 weeks is about 15% [6].

Many studies have reported the relationship between premature infants and disabilities. Nearly 8% of VLBW infants have disabling cerebral palsy (CP) at age 2 [9] [10]. 5% of VLBW infants have severe (moderate to profound) mental retardation at school age [11] [12]. The percentage of VLBW infants – who have learning difficulties, attention deficit disorder and strabismus - has been estimated to be as high as 40%, 30% and 15%, respectively [13].

Several studies have reported that 30% to 50 % of prematurely born children who display impaired academic achievements and/or behavioral disorders require additional educational resources [14]-[19]. However, the association between GM/IVH and developmental outcome in premature born children is still uncertain and controversial. It is unclear whether GM/IVH leads to children's developmental disabilities. Therefore, the topic is worthy of further investigation.

III. Definition of GMH/IVH

Germinal Matrix Hemorrhage (GMH):

Understanding the anatomy and neuropathology of germinal matrix is essential, because majority of the intraventricular hemorrhage originates from the germinal matrix [20]. Germinal matrix tissue is found mean the caudate nucleus [13] [20] directly ventrolateral to the lateral ventricle [20]. The germinal matrix is a temporary embryonic structure of the telencephalon existing before 36 weeks of gestation. It not only serves as the source of cerebral neuronal precursors from 10 to 20 weeks of gestation but also plays a role in the

formation of cerebral oligodendroglia and astrocytes. Cells for the pulvinar (dorsal and posterior portions) of the thalamus originate from the caudothalamic germinal matrix [21] that provides the neuronal cell clusters for cortical association regions to connect with the thalamic nuclei [22].

The germinal matrix from 28 to 32 weeks is noticeable on ultrasound (US) in the thalamostriate groove at the level of the head of the caudate nucleus at the site, of or lightly posterior to, the foramen of Monro [23]-[28]. That is the most common site for germinal matrix hemorrhage [20]. The germinal matrix is surrounded by, and is divided from, the cerebrospinal fluid (CSF) of the lateral ventricles by a single layer of ependyma. Once the ependymal lining breaks, intraventricular hemorrhage (IVH) occurs through extension of the bleeding into the ventricles.

The definition of germinal matrix hemorrhage is based on US diagnosis, and it is described as follows: on at least one ultrasound scan, there is a focal echodensity in the thalamocaudate groove that is just lateral to the frontal horns of the lateral ventricles. It sometimes extends to the head of the caudate nucleus [12].

The germinal matrix generally has involuted by the 32^{nd} to 34^{th} week of gestation [13] and is essentially exhausted by term [20]. That is the reason why the incidence of GMH among term neonates is rare but this type of intracranial hemorrhage is commonly seen in preterm infants [13].

Intraventricular hemorrhage (IVH)

In accordance with the above statement, intraventricular hemorrhage is a further pathological progress usually originating from germinal matrix hemorrhage. The most general source of intraventricular hemorrhage is rupture of the ependymal layer and extension of bleeding into the lateral ventricles [13]. It is difficult to ascertain the origin of bleeding in terms of the specific germinal matrix location [13].The following is a brief definition of IVH based on US diagnosis: on at least one ultrasound scan, there is an echodense focus or foci within the lateral, third or fourth ventricles separate from the choroid plexus, and it must be at least as echodense as the choroid plexus [12]. It can also be diagnosed when inequality of the choroid plexus margin reveals adherent intraventricular blood [12].

IV. Classification of IVH in the Premature Infant

Papile, Burstein, Burstein, and Koffler advanced the first classification strategy for intraventricular hemorrhage in 1978 [2]. Although Papile et al. used computed tomography (CT) scan to describe the results of brain imaging [2], their classification system provides the most common and widely used set of terms to describe ultrasound images of brain damage in preterm infants. Intraventricular hemorrhages have been classified into four separate grades in a hierarchical scheme numbered from I to IV based on the CT abnormalities to describe the "varied natural history of IVH" [2].

Grade I: Subependymal hemorrhage

Grade II: Intraventricular hemorrhage without ventricular dilatation

Grade III: Intraventricular hemorrhage with ventricular dilatation

Grade IV: Intraventricular hemorrhage with parenchymal hemorrhage

The classification of Papile and her coworkers is based on the location and amount of bleeding in an infant's brain. In addition, whether the hemorrhage enlarges or not is also important for the grades of severity of intraventricular hemorrhage.

Neonatal cranial ultrasound (US) scanning was introduced in the early 1980s, which enabled people to obtain more knowledge about the anatomy and neuropathology of an infant's brain. Volpe in 1987 provided a classification of IVH modified from that of Papile et al. In accordance with neuropathologic findings, Volpe suggested that the presence of parenchymal lesions (IVH grade IV in Papile's category) should be noted separately. That is because periventricular hemorrhagic infarction or other parenchymal lesions usually do not only result from extension of GM/ IVH into normal brain parenchyma. Therefore, there are three grades for IVH and a separate notation for periventricular echodensity in Volpe's classification [20]. The basis of Volpe's classification is the occurrence and proportion of blood in the germinal matrix and lateral ventricles [20].

Grade I: Germinal matrix hemorrhage with no or minimal Intraventricular hemorrhage (10% of ventricular area on parasagittal view)

Grade II: Intraventricular hemorrhage (10%-50% of ventricular area on parasagittal view)

Grade III: Intraventricular hemorrhage (>50% of ventricular area on parasagittal view;

usually distends lateral ventricle)

Separate notation: Periventricular echodensity (location and extent)

Although Papile's system is widely used, Paneth in 1999 indicated it has some weaknesses. First, it is cumulative. As a result, entities such as isolated ventricular enlargement or isolated parenchymal hemorrhage fall nowhere in Papile's system, because each grade that is above grade II includes the entities below it in the hierarchy [29]. Second, it lacks strong pathologic proof to indicate that grade IV hemorrhage is usually an extension of subependymal hemorrhage [29]. Preterm infants' parenchymal hemorrhages tend to be a component of white matter damage (WMD). For instance, bleeding into a pre-existing infarction, or microscopic hemorrhage presents together with periventricular leukomalacia (PVL) [29]. In addition, so-called "grade IV hemorrhage", which on US is an extensive echodense region in white matter surrounding the lateral ventricle, is sometimes not bleeding at all [29]. Many interpretations about "grade IV hemorrhage" based on ultrasound images are not appropriate for natural history of IVH [29]. Therefore, Paneth addressed the following three general categories involving not only GM/IVH but also the common brain lesions in premature infants. These categories are more concordant with neuropathologic discoveries.

1. WMD

2. Hemorrhages in non-parenchymal areas of the brain

3. Lesions in other brain locations: cerebellum, basal ganglia, brain stem, etc [29].

Whitaker (1996, 1997) et al. and Pinto-Martin (1999) et al. published studies that did not grade the severity of IVH in a hierarchical scheme numbered from I to IV. Instead, there were only two groups in their studies: (1) Isolated germinal matrix hemorrhage and/ or intraventricular hemorrhage (GM/IVH). (2) Parenchymal lesion and/ or ventricular enlargement (PL/VE) with or without GM/IVH [10]-[12]. The two groups correspond to grades I/ II IVH and grades III/ IV IVH respectively.

This thesis discusses the outcomes in premature children with germinal matrix and intraventricular hemorrhage (GM/IVH). "Grade III/ IV IVH" or "PL/VE" are not studied here. Therefore, the term "GM/IVH" will be utilized consistently in the thesis, which would not conflict with the different terms such as "grade I/II IVH" but specifically excludes "grade III/ IV IVH" and "PL/VE" as described above.

V. Diagnosis of IVH

Ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) are the primary methods currently used in the evaluation of premature infants' brains [30]. CT has been available for the past three decades [30]. US scan has been widely used since its introduction in the early 1980s [10][14]. Although MRI has been shown to provide superior images of IVH, particularly after the first few days of brain hemorrhage [31]-[34], it currently cannot supplant US in the evaluation of IVH [20]. This is because MRI has some disadvantages: (1) it requires transporting the infant to the scanner, (2) it has a relatively long data acquisition time, (3) it prevents use of metallic materials found on standard neonatal

monitoring and support equipment, and (4) it is expensive [20]. Therefore, the following discussions exclude MRI and only focus on US and CT whose principles, strengths, and weaknesses are described.

Computed Tomography (CT)

Computed tomography (CT) is sometimes called CAT scan. Special x-ray equipment is used to obtain many images from different angles and then these images are combined to show a cross-section of body tissues and organs [30]. CT scan is able to demonstrate the site and extent of intraventricular hemorrhage effectively [2] [4] [20] [35]-[42].

CT scan provides a poorer image of the precise details of soft tissues (particularly the brain) compared with MRI [30]. On the other hand, it is more sensitive than ultrasonography in cases where extra axial (subdural, subarachnoid) or posterior fossa haemorrhage is suspected [20]. However, CT scan has major disadvantages: (1) the infant is required to be removed from the intensive care unit (ICU), and (2) the brain and eyes of the infant will be exposed to ionizing radiation [20] [30]. In addition, it is not a portable technology and is thus gradually replaced by US scan [20].

Ultrasound (US) Brain Scan

Ultrasound scan has high reliability and versatility in identifying the degree of severity of IVH, which ranges from isolated germinal matrix hemorrhage to major degrees, with or without periventricular hemorrhagic infarction [20]. Most cranial ultrasound studies in neonates utilized a sector or linear-array transducer with a frequency of 7 MHz or higher

[43]. Images of ultrasound scans are collected through the anterior fontanelle, in both the coronal and sagittal planes [13] [20]. The posterior fossa and midbrain can be imaged well through the posterior fontanelle [13] [20].

Performing three US scan at 4 hours, 24 hours, and 7 days of life can improve the sensitivity of US detection to more than 75% for GM/IVH [13]. The screening examination at 7-14 days of life can detect more than 80% of GM/IVH which may result in post-hemorrhagic hydrocephalus [20] [30] [43]. Furthermore, the sensitivity of ultrasound for detecting white matter abnormalities can be increased further by performing weekly scans after the first month of life [43]. Later scans at 36 to 40 weeks postmenstrual age are recommended by the Committee of the American Academy of Neurology and the Child Neurology Society to evaluate infant prognosis [30].

The major advantage of US is its portability [30]. This is because US can be performed without moving infants from the intensive care unit (ICU) [30]. In addition, it is safer and cheaper than CT and MRI [20].

Chapter Two
Methods of this Research

I. Hypothesis

This study has a null hypothesis: there is no significant difference for premature children's neurodevelopmental and cognitive outcomes between the group of infants who have GM/IVH and those who with normal results of ultrasound scanning. On the other hand, an alternative hypothesis of this study is that there are significant differences for those outcomes between GMH/IVH group and normal ultrasound scanning group. I here extract evidence from the literature to evaluate and ascertain this hypothesis by systematic literature review in this thesis.

II. Literature Search

1. Sources for the Literature Search

(i) Bibliographic Databases on the Internet

Search engines or bibliographic databases on the Internet are the major sources of literature that will be analyzed in this thesis. These sources store most of the current research and are convenient for literature search. These sources include MEDLINE/ PubMed, ISI Web of Knowledge^{SM}, electronic journals, and a powerful searching engine "Scholar Google".

MEDLINE/ PubMed— MEDLINE is a comprehensive and cross-referenced database of citations provided by the National Library of Medicine (NLM) [44]. It includes medical

literature from 1966 to the present. Searching MEDLINE through PubMed provides researchers several advantages such as a user-friendly interface, cosmopolitan search resources, and access to full-text articles [45]. The following components are included in the MEDLINE database: Title, Author(s), Affiliation, Abstract, Language, Publication Date, Journal Title, and Medical Subject Heading (MeSH) terms [46]. NLM defines the MeSH terms chosen from 22,568 descriptors and continually updates them [46].

The PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed) provides a typical single search window (Figure 1). By entering keywords in that window, a link to the full-text of more than 4,400 journals will be provided by the PubMed searches [47]. Results of the literature searches from PubMed include literature (1) with free full texts, (2) with abstracts but without free full texts, or (3) without abstracts and free full texts [44]. For literature with abstracts but without free full texts, usually only subscribers of those journals can access the full text. However, non-subscribers may still access those articles by paying a fee [44]. Once a useful literature is found but cannot be downloaded from PubMed, other sources and search methods will need to be used.

ISI Web of Knowledge^{SM}— ISI Web of Knowledge^{SM} is another electronic academic literature source, which is as important as MEDLINE/ PubMed for this thesis. It enhances the function of ISI Web of Science and provides three major databases: "Science Citation Index Expanded", "Social Sciences Citation Index" and "Arts & Humanities Citation Index". For this thesis, we only use Science Citation Index Expanded for literature search in ISI Web

of Knowledge℠, because it provides articles that belong to various science fields. Through ISI Web of Knowledge℠, it is possible to search for articles from more than 8,500 journals in different fields of science. In addition, it has a unique function, called "Cited Reference Searching", which provides an easy pathway of cross citation.

However, it is not convenient for individuals because the use of "ISI Web of Knowledge℠" needs to be licensed through academic registration. Therefore, people with no "ID" and "Password" from an academic register cannot access "ISI Web of Knowledge℠". For this thesis, literature searches from ISI Web of Knowledge℠ are administered by registering via the Proxy Server of Michigan State University Libraries.

Electronic Journals— It is possible that people obtain only abstracts without full-text articles while searching for articles on MEDLINE/ PubMed or ISI Web of Knowledge℠. Under this circumstance, searching for literature from electronic journals is an appropriate and practical solution. The results of literature searches from MEDLINE/ PubMed or ISI Web of Knowledge℠ do not always provide the address of a journal's website. Therefore, the best way to search a journal's website is through the libraries' networks of academic institutions, such as universities, colleges, or research centers. In this thesis, I access individual journal's website through the homepage of Michigan State University Libraries (http://www.lib.msu.edu/).

Scholar Google— Scholar Google (http://scholar.google.com/) is a very powerful and useful search engine similar to Google (http://www.google.com/). The difference between

Scholar Google and Google is that Scholar Google focuses only on academic papers and books but not on other irrelevant links. Obtaining the knack of how to effectively and efficiently utilize this public search engine is crucial for literature retrieval [48]. The main advantage of using Scholar Google for literature search is that it not only combines several academic databases but also provides links of citations. Scholar Google is a valuable source for literature search in this thesis because it can find literature that cannot be found from PubMed or ISI Web of Knowledge^{SM}.

(ii) Medical or Epidemiological Journals in Libraries

The medical or epidemiological journals stored in libraries are other important literature sources for this thesis. Databases on the Internet are convenient sources for searching recent publications. For epidemiological observational studies, however, only some of the relevant articles can be found by a computer-aided literature search [49]. The reason is that academic websites do not always provide the complete databases, and the searches may not return all relevant literature. In addition, some of the full-text literature published a long time ago may not be in the databases. Therefore, printed journals in libraries still have a great value and cannot be replaced or ignored. In this thesis, some articles are from printed journals such as *Journal of Developmental and Behavioral Pediatrics*, *Pediatrics*, *Archives of General Psychiatry*, *Seminars in Perinatology*, *Developmental Medicine and Child Neurology*, *Journal of Pediatric Ophthalmology and Strabismus*, *the Journal of Pediatrics* and *Early Human Development*. Most of these journals were found in libraries of

Michigan State University (MSU) and University of Michigan at Ann Arbor. If a journal could not be found in these two libraries, I asked the librarian to recall the journal from other libraries, or I recalled the journal on the website of MSU library by myself.

2. Determining Key Concepts and Generating Search Terms

The research question of this thesis is: "Is GM/IVH causally associated with neurodevelopmental and cognitive abnormalities in premature children?" Breaking down the research question into some key concepts and then generating useful search terms is the first step to performing a literature search [46] [50].

(1) Concepts about exposure include the following terms: germinal matrix hemorrhage, subependymal hemorrhage, intraventricular hemorrhage, periventricular hemorrhage, periventricular-intraventricular hemorrhage, mild brain damage, and intracranial hemorrhage.

(2) Concepts about other risk factors include the following terms: low birth weight, low-birth-weight, low birthweight, preterm, preterm infant, premature and, prematurity.

(3) Concepts about outcomes include the following terms: neurodevelopment, neurodevelopmental, neurodevelopmental outcome, neurodevelopmental performance, cognition, cognitive, cognitive outcome, cognitive performance, learning, behavior, outcome, and school performance.

To perform the literature search effectively, I chose two or more terms from different concept groups and then typed them in the search box. The results are shown after clicking

the "Go" button on the PubMed web page. MeSH terms lead retrievals into limited records on that specific topic [51] and may improve the specificity of literature searches [52]. On the other hand, a free-text search, which is searching through generic words, retrieves records including the given word anywhere in the title or abstract even if these records are not essentially related to the subject [53]. To obtain optimal results, therefore, free-text search needs to be performed with a MeSH search [54].

Finally, it is very important to note that English is the sole language adopted by MEDLINE field contents [46]. Hence, all keywords have to be in English for any literature search through MEDLINE.

3. Strategies for Performing MEDLINE/ PubMed

MEDLINE allows Boolean operators [55] to combine keywords for the searches [46]. AND, OR, and NOT are the connectors which may improve search specificity [53] [55]. All the connectors have to be capital letters in MEDLINE [55] [56]. To retrieve a set of records simultaneously covering all given search terms, the AND operator should be utilized [46]. On the other hand, records contain any of the given terms can be retrieved by the OR operator [46]. By using NOT operator, only records that involve the first key word will be returned, and those related to the second will be excluded [46]. In addition, utilizing the truncation function (*) can increase the search recall due to synonyms (e.g., hemorrhage and bleeding) and variants (e.g., behaviour and behavior) of a term [46]. For a multiword keyword, typing the keyword in quotation marks (e.g., "low birth weight") is appropriate

and beneficial [46]. Following these strategies may help researchers to conduct efficient and effective literature search.

III. Literature Selection

1. Inclusion and Exclusion Criteria

Any review paper should have explicit inclusion and exclusion criteria in order to appropriately select all relevant studies [56] [57]. The criteria of this review are listed as follows.

(1) Only published papers written in English are included. Published papers written in other languages, such as Chinese or Spanish, are excluded.

(2) This thesis includes only relevant papers published after 1980 and excludes papers published before 1980.

(3) The included studies should involve GM/IVH (IVH Grade I/II) and have a clear classification of IVH. Studies are excluded if the authors only address PL/VE (IVH Grade III/IV) or combine GM/IVH with other types of brain damage (e.g. PL/VE).

(4) Included studies must involve preterm (gestational age $\leqq$ 37 weeks) or low birth weight (birthweight< 2.5 kilograms) subjects.

(5) Studies included in this thesis must have a premature normal-ultrasound group to compare with the premature GM/IVH group.

(6) All the included studies should be related to neurodevelopment, mental development, school performance, behavior, or cognition.

(7) Sample size is an important criterion for selecting literature. The sample size of each study group in the included papers has to be larger than 10.

(8) The age of assessment has to be equal to or older than 10 months and equal to or younger than 18 years old.

(9) All the included studies should provide the number of surviving infants.

2. Selection with Titles, Abstracts, and Texts of Literature

The next step for literature selection is scanning titles and abstracts once an acceptable number of abstracts have been found [52]. During the literature selection of observational studies, investigators may find that the titles of the publications usually have clear descriptions of the subjects, exposure, outcome, or research method. The information provided by those titles should be examined through the inclusion and exclusion criteria. It is really a quick and initial measure for literature selection. However, the major misgiving during the scan of study titles is that authors and selectors may differ in how they define concepts [58]. These differences in defining concepts are important causes for failed database searches and selection [59]. A literature selection relying on keywords in the titles of studies will result in the neglect of many relevant publications [58]. For instance, some publications related to the topic of this thesis have different keywords for the ultrasound scan results in their titles, such as "intraventricular haemorrhage", "periventricular-intraventricular haemorrhage", or "cranial ultrasound abnormalities". If a study selector has different definitions of those key terms, many useful relevant publications

may be missed. Therefore, good literatures may not be selected if only the titles of the papers are scanned. Scans of abstracts are also important and necessary.

Structured abstracts usually consist of one or more of the following components to assist in the selection of relevant publication: background, objective, design, methods, results, and conclusion. Nevertheless, few abstracts address all the essential information, and the study method or design is sometimes missing [58]. Therefore, a researcher should select literature not only through scanning titles and abstracts but also through carefully reading the texts of publications.

IV. Literature Analysis

1. Determining the Type of Literature Review

Literature reviews can be classified into at least four types: (1) qualitative and narrative reviews, (2) quantitative reviews of published data (usually called meta-analysis), (3) meta-analyses with individual data (usually called pooled analysis), and (4) prospectively planned pooled analyses [57]. Each type of literature review has its own strengths and limitations. Clearly evaluating those strengths and limitations is an inevitable stage in determining the type of literature review that is appropriate for the research in this thesis.

Traditional narrative reviews only present qualitative results [57]. Publication bias [60][61], often called "file-drawer" problems [62], is the main limitation of traditional narrative reviews. The "file-drawer" problems arise because the results of unpublished literature are imagined to be guzzled in investigators' file cabinets [62] [95]. To avoid

subjective judgments of the selected studies, an a-priori strict protocol for the review is absolutely necessary. Following the indication of the explicit protocol, an extensive overview of the included studies can be done within a short period of time and at a low cost [57].

Meta-analysis from published data has several main limitations. First, publication bias cannot be avoided because some epidemiological researches may be adventurously done, and the positive results are more likely to be published. Therefore it is possible that meta-analysis provides an overestimation of the risk estimate [57]. Second, the included studies may differ in their design, assessment of exposure or outcome, and definition of confounder variables. Adjusting for different confounders in separate studies is very difficult for meta-analysis [57]. In addition, reliable results cannot be derived from meta-analysis if the heterogeneity among included studies is high [57].

Meta-analysis with individual data can avoid some of the problems that arise with meta-analysis from published data. This type of analysis may eliminate publication bias because it includes unpublished data. With individual data, investigators can conduct a statistical re-analysis and reduce some errors in the analysis [57]. Furthermore, it is feasible for investigators to examine the effect of rare exposures with a large sample size from several separate studies [57]. However, this type of literature review has its own limitations. It takes a lot of time and money to do the analysis [63]-[65]. In addition, the direct cooperation between study coordinators is necessary [63]-[65].

Prospectively planned pooled analysis requires a joint core protocol for data collection and analysis between separate studies to eliminate their large heterogeneity [57]. However, it is very time-consuming and difficult to perform. In addition, prospectively planned pooled analysis may multiply errors in the design of single studies [57].

The literature review method used in this thesis does not belong to either meta-analysis review or prospectively planned pooled analysis review. The main reason is that the heterogeneity among the included studies is very high. Although most of the included publications use cohort study design and adopt ultrasound scan to measure the exposure, the assessed age and the measurement of outcomes among those studies are highly different. Therefore, it is inappropriate to classify this literature review method as meta-analysis. In addition, this thesis presents only the published data, and it excludes unpublished data or primary data from all included studies. For this thesis, it is impractical to obtain the original data from the coordinators of the included studies due to the lack of financial support, scanty manpower, and limited time.

In the thesis, published data from the included studies were carefully organized and synthesized. The relevant components of each study, including authors, publication year, study design, sample size and assessment procedures, are carefully tabulated. Procedures for summarizing the Evidence among studies include: (1) systematic organization and clear presentation for the statistical data, and (2) objective judgment and logical explanation for the results. Methodologic guidelines were adopted for this literature review [56][57].

By generalizing the above statements, as well as considering the materials, characteristics, and sources of included literature in this thesis, I decided to use a review strategy that combines qualitative and quantitative methods and called this a "systematic" review.

2. Methods for Summarizing Evidence

A review published by Breslow et al. indicated that having clear methods for summarizing evidence among studies is essential [66]. Pooling of odds ratios, relative risks, and methodologic quality ranks are some of the quantitative methods adopted by most of the review papers examined by Breslow et al. [66]. I observe that the validity of individual studies can be appropriately assessed through the study-quality evaluation process. Therefore, I apply two methods in this thesis to obtain high-quality literature summaries, including evaluations using causal criteria and study-quality criteria.

(i) Evaluation with Causal Criteria

Two series of review papers, which were reviewed by Weed et al. [67], evaluated the study of causation using causal criteria [68]. Most of those review papers used Hill's causal criteria [69]. Those criteria provided a practical method to judge whether causation, which is based on the evidence that appeared in research, exists [70]- [72]. Therefore in this thesis we adopt and administer an evaluation using Hill's causal criteria for literature analysis.

Hill's nine causal criteria are strength, consistency, specificity, temporality, biological

plausibility, biological gradient (dose-response), experiment, coherence and analogy [69]. Parts of the nine criteria are established through the review process and are used to evaluate the 16 selected studies in order to approach the possible existent causation.

Strength— All studies selected in the thesis adopt a cohort study design. For cohort studies, strength is a useful criterion to evaluate the extent of association between exposure and outcome. The association is achieved by comparing the outcome of an exposure group with that of a non-exposure group. Strength can also be used to examine the association in case-control studies by comparing the exposure of a disease group with that of a non-disease group.

Causal inferences are more likely to be derived from strong associations than from weak ones [73] [74]. However, it does not mean that weak associations cannot support the establishment of causal association. Sometimes the strength of association is small because of the effects from other risk factors or bias. Therefore, third variables and bias have to be seriously estimated when researchers find weak associations in their studies.

Consistency— This criterion is used to evaluate whether the association has been repeatedly observed by different persons, in different places, at different times, and under different circumstances [69] [73].

Specificity— This criterion means that a specific exposure results in a specific outcome. If specificity is established, stronger case for causation may be made [73]. In this thesis, however, specificity cannot be established because the included studies have various

outcomes, even if they have the same exposure (GM/IVH).

Temporality— This criterion is established in all selected studies because they all adopted prospective cohort study design. In addition, GM/IVH always occurs before children's developmental conditions can be examined. Since temporality is established in every selected study, this criterion is not useful for judging the causation in this thesis.

Biological plausibility— This criterion means that there is a coincidence between the observed associations and current biological knowledge [69] [73] [74]. It is useful to consider biological plausibility because biases or confounding factors may be easily identified by this criterion [74]. Many clinical or bio-medical studies use biological markers to establish biological plausibility. In this thesis, however, no study adopted biological markers or similar relevant methods to establish the biological plausibility.

Biological gradient (dose-response)— This thesis discusses the causal association between GM/IVH and children's neurodevelopmental and cognitive outcomes. There is no biological gradient established in this thesis since the results of more severe brain hemorrhage are not included and discussed.

Experiment— It is impossible to split subjects into different groups and then expose them to a risk (GM/IVH) or non-risk (no hemorrhage) condition. This criterion is not used, because all included studies use a non-experimental design.

Coherence— Coherence means there are no irreconcilable conflicts between causal associations derived from study data and general epidemiological, medical, or biological

knowledge [69] [74]. Most included studies comply with this criterion.

Analogy— The criterion is not used in this review because of limited evidence with other kinds of subjects (e.g. assessing the outcomes of adults', elders', or rats' brain bleeding).

In this thesis, a criterion that is established in all or none of the studies is not useful for judging causation. As a result, it is not useful to adopt a criterion established in all selected studies such as "temporality"; similarly, it is not useful to adopt criteria that are not established in any of the selected studies, such as "specificity", "biological plausibility", "biological gradient", "experiment", and "analogy". Since only the criteria that are established in part of the selected studies can show the difference of causation, I use strength, consistency and coherence to evaluate the causation in this research.

(ii) Criteria for Evaluating the Quality of Each Study

Setting criteria for study-quality evaluation [56] is essential for systematic literature reviews. In this thesis, we first consider several aspects of a study and then provide a study quality checklist [52].

The criteria used in this thesis to evaluate study quality include: (1) whether the follow-up rate is equal to or larger than 90%, (2) whether the assessment rate is equal to or larger than 85%, (3) whether the mean of gestational age between the study groups is not significantly different, (4) whether the mean of birth weight between the study groups is not significantly different, (5) whether US brain scan was used for GM/IVH diagnosis, (6)

whether the study has a full-term control group, and whether the full-term control group is well-matched, (7) whether the sample size is equal to or larger than 30, (8) whether it has careful or standard measurements of outcomes, (9) whether key confounders are mentioned and adjusted, (10) whether the other risk factors are mentioned and controlled, and (11) whether it has a clear conclusion about GM/IVH.

Based on the above criteria, a study will be given one point if its answer for a criterion is "yes". On the other hand, it will be given zero point if its answer for a criterion is "no". The more points a study receives, the higher study quality it has.

In addition, the age of assessment is the twelfth criterion. A study with an older age of assessment is regarded as a higher quality study. The age of assessment is divided into five ranges, and the score for each range is determined as follows: (1) $\geqq$10 months & $\leqq$1.5 years: 1 point; (2) >1.5 years & $\leqq$3 years: 2 points; (3) >3 years & $\leqq$5 years: 3 points; (4) >5 years & $\leqq$10 years: 4 points; (5) >10 years & $\leqq$18 years: 5 points.

Chapter Three

Results

I. Studies Identified

Two hundred to three hundred publications were reviewed. By verifying these publications against the inclusion and exclusion criteria, almost 90% of them were excluded. Initially, twenty-three publications were identified to be useful. However, some of them did not meet all the inclusion criteria: three of these publications combined GM/IVH with other types of brain damage and had no clear IVH classification; two of these publications assessed the subjects at the age of three months; one publication had no premature normal ultrasound group; one publication did not provide the size of the birth cohort. Therefore, seven of the twenty-three publications were excluded. Finally, sixteen journal articles were selected for this thesis.

II. Introduction of Birth Cohorts

All included studies adopted the cohort study design. Tables 1-A and 1-B present the essential information about the birth cohort of each included study. The information includes: study location, number of birth cohort members, admission criterion (i.e. gestational age and birth weight), average of gestational age and average of birth weight in each group. Table 1-B provides the information on the birth cohorts with both a premature group and a full-term control group, while Table 1-A includes studies with only premature groups. Three studies conducted by van de Bor et al. (1988, 1993, and 2004) were from

the same birth cohort (Table 1-A). Two studies conducted by Ross et al. (1992; 1996) were from the same birth cohort (Table 1-B). In addition, studies conducted by Pinto-Martin (1999) are of the same birth cohort as the study conducted by Whitaker et al. (1996; 1997) (Table 1-A). Therefore, eleven birth cohorts are represented in the sixteen selected studies.

Ten of the studies were conducted in six different cities or counties in the United States [10] [12] [14] [75]-[81]. Of the other studies, three were conducted in the Netherlands [1] [82] [83], two in Australia [84] [85], and one in the United Kingdom [86].

The gestational age of the premature birth cohort is not uniform but is generally less than 34 weeks for all included studies. However, nine studies did not provide the range of gestational age for recruiting the cohort [10] [12] [14] [75] [76] [78] [81] [85] [86]. Table 1-A shows that gestational age is significantly different between the GM/IVH group and the premature normal ultrasound group in eight studies [1] [10] [12] [14] [76] [78] [82] [83]. Overall, the mean gestational age of the GM/IVH groups of all studies is 29.6 weeks (SD= 1.9). The mean gestational age of the premature normal-ultrasound groups of all studies is 30.7 weeks (SD= 1.9). The mean gestational age of subjects with GM/IVH is significantly different than those with normal ultrasound (P= 0.01). This means that gestational age may be a confounding factor for the relationship between GM/IVH and children's developmental outcomes. In addition, the mean gestational age of the full-term control group of all studies is 39.4 weeks (SD= 1.1).

Birth weight is classified into three ranges: low birth weight (LBW, <2500 grams), very

low birth weight (VLBW, <1500 grams) and extremely very low birth weight (EVLBW, <1000 grams). Four studies involved the LBW cohort [10] [12] [14] [77], nine studies the VLBW cohort [1] [75] [76] [78] [81]-[83] [85] [86] and one study the EVLBW cohort [84]. Two studies did not provide the enrolling criteria of birth weight for their cohort [79][80]. From Table 1-A, I observe that in four studies [10][12][14][78], the birth weight is significantly different between the GM/IVH group and the premature normal ultrasound group. Overall, the mean birth weight of the GM/IVH groups of all studies is 1300.5 grams (SD= 274.2). The mean birth weight of the premature normal ultrasound groups of all studies is 1393.1 grams (SD= 282.5). The mean birth weight of subjects with GM/IVH is not significantly different than those with normal ultrasound (P= 0.08). Furthermore, the mean birth weight of the full-term control group of all studies is 3719.3 (SD= 349.7).

III. Measures and Classification of IVH

Appendix Table 1 summarizes the measures used and classification system of IVH in the studies. From this table, I observe that the measurement of IVH tends to be consistent among the included studies: fifteen studies utilized ultrasound scans to examine the infant's brain pathological status. But, two studies published by Papile et al. (1983) and Ment et al. (1985) adopted both ultrasound scans and computed tomography brain scans to examine their subjects, while one study, reported by Lowe et al. (1990), utilized only computed tomography brain scans to measure the subjects' IVH status.

Times and timeframe of IVH measurements among the included studies are not

uniform but are very similar. IVH status was examined at least two times and at most four times. The initial examination was conducted right after birth or right after the subjects were admitted into the intensive care unit of hospitals. The timeframe of the last examination was slightly different among studies; however, it was never longer than the first two months of life.

Various terms and abbreviations were adopted by the included studies for the classification of IVH. Eight of the studies directly used Papile's system for classifying the IVH [1] [75] [76] [78] [81]-[84]. The other studies utilized a modified classification to assign the subjects into the study groups. The IVH classifications are all compared in Appendix Table 1. In general, "Grade I/II IVH", "Grade I/II CIVH", "Grade I/II PIVH", "Low-grade hemorrhages", and "S/IVH" are uniformly called "GM/IVH" in this thesis.

IV. Follow-up and Assessment Rates in Studies

Tables 2-A and 2-B provide important statistical information about assessed children: the age of assessment, the follow-up rate, the assessment rate, and the number of children who successfully completed the developmental tests (i.e. number of assessment). Table 2-B summarizes information about studies that assessed children with a premature groups and a full-term control group. In Table 2-A, studies include only premature groups.

The neurodevelopmental and cognitive outcomes were assessed at different ages in the selected studies. Five studies had one or more ages of assessment [10] [75]-[78]. The ages of assessments were classified into five ranges in this thesis to analyze the potential

heterogeneity: (1) 10 months to 18 months, (2) 18 months to 3 years, (3) 3 years to 5 years, (4) 5 years to 10 years, and (5) 10 years to 18 years. Five studies were in the range from 10 months to 18 months [75]-[78] [80]. Seven studies were in the range from 18 months to 3 years [10] [75]-[79] [82]. The range from 3 years to 5 years included three studies [83] [85] [86]. The range from 5 years to 10 years included five studies [10] [12] [14] [81] [84]. There was one study in the range from 10 to 18 years [1]. Overall, the mean age of assessment of all studies is 4.5 years which is larger than 3 years and is sufficient for high study quality.

Follow-up rate and assessment rate of the selected studies were extremely different, ranging from19% [81] to 100% [80]. This thesis emphasizes the follow-up rates of the premature groups, and does not focus on those of full-term control groups. This is because most studies with a full-term control group did not recruit the full-term subjects at birth, except the study published by Sherlock et al (2005).

Follow-up rate represents the percentage of children who were still alive in the age of assessment and were contacted successfully. Assessment rate means the percentage of follow-up children who accepted the given developmental tests and could complete the tests successfully. Therefore, the assessment rate is usually lower than the follow-up rate. Higher follow-up and assessment rates are better. In this thesis, if a selected study has a follow-up rate higher than 90% and has an assessment rate higher than 85%, it is regarded as a study with high quality. Based on Tables 2-A and 2-B, eight studies reported respectively

by Kitchen et al. (1990), van de Bor et al. (1988, 1993 , and 2004), Levene et al. (1992), Sherlock et al. (2005), and Ross et al. (1992; 1996) met this criterion.

Overall, the mean follow-up rate and the mean assessment rate among studies are 83% and 75%, respectively. Although the mean follow-up rate and the mean assessment rate do not meet the criteria of high study quality, they are still acceptable and can provide some evidence for this research.

In addition, "number of assessments" provides a clear picture of the composition of study groups that indicate the number of children who successfully completed the various development tests. Table 2-B shows that Lowe et al. (1990) and Ross et al. (1992; 1996) conducted studies focused on GM/IVH but did not focus on PL/VE.

V. Study Quality Evaluation

In this review, the main quality variation of the included publications is caused by difference in follow-up rates, assessment rates, sample sizes of study groups, age of assessment, and adjustments for key confounders and other risk factors. Any variation in one or more factors affects the quality of the included publications.

A study whose follow-up rate and assessment rate are equal to or higher than 90% and 85% respectively is recognized as a high quality study. Table 3-A shows that half of the 16 included studies meet the former criterion. On the other hand, 10 studies meet the later criterion; in other words, that the assessment rate is equal to or higher than 85%.

Birth weight and gestational age, combined with cranial ultrasound findings, are the

most sensitive predictors of sensory, cognition, language, psychomotor and academic performance in early school-aged children as well as in younger children [87]. As a result, if there is a significant difference in mean birth weight and gestational age between premature children with GM/IVH and those without GM/IVH, the birth weight and the gestational age may confound children's developmental outcomes. Therefore, a study is regarded as high quality if its enrolled infants, with and without GM/IVH, had no significant differences in mean birth weight and mean gestational age. Based on Table 3-A, I observe that seven studies had no significant difference in mean gestational age and twelve studies had no significant difference in mean birth weight.

Ultrasound brain scan is a reliable and standard technology for examining the presence of brain hemorrhage. The study reported by Lowe et al. did not utilize ultrasound for brain screening and thus does not meet the criterion of high-quality study (Table 3-A).

Five studies reported by Lowe et al. (1990), Levene et al. (1992), Sherlock et al. (2005), and Ross et al. (1992; 1996) had matched full-term control groups (Table 3-A); therefore, they are regarded as high-quality studies. Those full-term control groups matched the premature groups in different dimensions such as race, age, gender, and parents' socio-economic status (Table 1-B). Through a careful match, the effects of those variables can be controlled. Although premature normal ultrasound group already plays the role of control group, a study with a matched full-term control group still provides additional valuable advantages. With a full-term control group, the effects of birth weight and

gestational age on outcomes can be investigated. On the other hand, studies without a full-term control group can only investigate the effects of GM/IVH on the outcomes. In addition, the external validity of a research can be increased by the comparison between the full-term control group and the premature group.

A larger sample size can lead to a more robust result. Based on the Central Limit Theorem [88], a distribution with a larger sample size is closer to a normal distribution. Usually, a sample size larger than 30 is assumed to result in an approximately normal sampling distribution. Table 3-A shows that three studies reported by Sostek et al. [77], Lowe et al. [81] and Ross et al. [79] do not meet the criterion of high study quality because some of the groups in those studies have sample sizes smaller than 30 (Tables 2-A and 2-B). If a sample size is very small, confidence interval will become so large that the Central Limit Theorem cannot be applied. In this case, no difference will be found between the GM/IVH group and the normal ultrasound group. Ross at al. obtained some significant findings when they administered the special cognitive tests (i.e. invisible displacement tasks and object discrimination reversal tasks), but Sostek et al. and Lowe et al. did not find any significant results. This might be due to their different sample sizes of groups: the group sizes in studies of Ross et al. are close to 30 (N=27 and 28), but those in studies of Sostek et al. (N=23 and 20) and Lowe et al. (N=22 and 11) are much smaller than 30. Obviously, further evaluation of the results from these studies is necessary because of their small sample sizes.

Several different measures were administered in the sixteen included studies, reporting a variety of developmental outcomes (Appendix Table 2). Attempting to design related studies with a uniform measurement is often challenging because human development and cognition are complex and comprehensive processes. As a result, they cannot be fully studied by evaluating any single outcome. Therefore, in this review, any study administering careful or standard measurements of outcomes is recognized as a high-quality study. All selected studies met this criterion (Table 3-B).

To explore the long-term neurodevelopmental and cognitive outcomes, a study with the age of assessment larger than 3 years is regarded as a high-quality study in this review. Three points were given if a study had the age of assessment between 3 and 5 years. Four points were given if a study had the age of assessment between 5 and 10 years. In addition, five points were given if a study had the age of assessment between 5 and 18 years. Table 3-B shows that 9 studies receive at least three points due to the criterion of the age of assessment.

Other risk factors and key confounders have to be either controlled while the researchers design a study or adjusted while the researchers do the statistical analysis [89]. Confounding may cause errors when estimating the association between a typical exposure (e.g. GM/IVH) and its outcome (e.g. cerebral palsy) [89]. Other risk factors, which may be precursors or mediators, also need to be recognized in order to estimate their influence on the outcomes. Six studies in Table 3-B identified and adjusted for the key confounders

such as gestational age, birth weight, Apgar score at 5 minutes < 7, small for gestational age, respiratory distress syndrome, seizures of all origins, apnea (more than 15 sec. and/ or bradycardia), hyperbilirubinemia (maximum neonatal serum total bilirubin > 200 μ mol/l) and bronchopulmonary dysplasia. On the other hand, nine studies mentioned and adjusted other risk factors such as social disadvantage, gender, outborn, assisted ventilation and multiple gestations (Table 3-B). These studies meet the study quality criteria.

Table 3-B reveals that four selected studies emphasized the outcomes caused by severe brain damage (e.g. ventricular enlargement and/or parenchymal lesion) and did not make a clear conclusion for the effects of GM/IVH. Therefore, these studies did not meet the criterion of high study quality.

VI. Ranking of Importance for Selected Studies

Every study was evaluated by the study quality criteria (Tables 3-A, 3-B). When a study met any of the first eleven criteria, it would obtain one point. In addition, a study with an older age of assessment would receive more points. After the evaluation, a study receives a total score that represents its quality (Table 3-B). Table 4 is created to provide the ranking of importance for selected studies by combining the evaluation of causal inference with that of study quality. The study of van de Bor et al. (2004), which is on the first place, is the most important study. On the other hand, the study of Sostek et al. (1987), is on the last place, showing that it is the least important study (Table 4). Studies of Kitchen et al. (1990), Levene et al. (1992), Sherlock et al. (2005), and Ross et al. (1992;

1996) tied as fourth place winners (Table 4). Eight studies, listed from the first to the fourth places, were regarded as important and reliable studies.

VII. Frequencies and Odds Ratios of Developmental Abnormalities

In Tables 5-A and 5-B, the data used to calculate the frequencies of abnormal or special cognitive outcomes were abstracted from twelve of the included studies, and the percentages of outcomes were calculated manually. In Table 5-A, the frequencies of abnormal outcomes in the GM/IVH groups and the normal ultrasound groups are compared. Table 5-B compares the GM/IVH groups with not only the normal ultrasound groups but also the full-term control groups.

In Table 5-A, eight studies [1] [12] [78] [81]-[85] had one or more abnormal outcomes with higher frequencies in the GM/IVH groups than in the normal ultrasound groups. The frequencies of abnormal outcomes in the GM/IVH groups were at least twice as high than in the normal ultrasound groups in the eight studies. Take the study reported by Kitchen (1990) for instance, the prevalence rate of cerebral palsy (CP) in the GM/IVH group is 18.18%, and in the normal ultrasound group is 3.13% (Table 8-A). As a result, the CP rate in the GM/IVH group is six times higher than in the normal ultrasound group. Performing the same comparison for the study published by Ross et al. in Table 5-B, I observe that the abnormal rate of children who habituated in the GM/IVH group (46.67%) is significantly lower than the normal ultrasound group (96.67%), as well as the full-term control group

(90.00%). In addition, the frequency of failed reversal trials in the GM/IVH group (73.33%) is significantly higher than not only the normal ultrasound group (56.67%) but also the full-term control group (16.67%). Based on the comparison of frequencies, I may infer that there might be significant differences in the outcomes between the GM/IVH groups and the normal ultrasound groups.

Odds ratios (OR), 95% confidence intervals (CI) and p-values were provided by the authors in five studies [1] [12] [14] [82] [83]. In the rest of the studies, those statistic results were calculated using the SAS program. Significant differences between the GM/IVH groups and the normal ultrasound groups existed in six studies [1] [12] [78] [82] [83] [85] for twelve abnormal outcomes in Table 5-A. Children with GM/IVH had more disability than those with normal ultrasound in two studies [83] [85]. In 1988 and 2004, two studies reported by Van de Bor et al. showed that children with GM/IVH had more handicap than those without GM/IVH at the age of 2 (OR=2.10; 95%CI=1.30-3.30; P<0.01) and 5 (OR=2.70; 95%CI=1.21-6.08; P=0.01), respectively. Based on the odds ratio and 95% CI in the study of Kitchen et al., cerebral palsy prevalence (OR=6.89; 95%CI=1.62-29.39; P=0.003) was regarded as significantly different between the GM/IVH group and the normal ultrasound group. Mental retardation (OR=5.00; 95%CI=1.50-16.80; P<0.01) was significantly elevated by GM/IVH in the study of Whitaker et al. (1996), and abnormal neurodevelopment (OR=4.90; 95%CI=1.96-12.27; P=0.0003) was significantly elevated by GM/IVH in the study of Hanigan et al. (1991). In the study of van de Bor et al.,

which was reported in 2004, the risk that children needed special education at 5 (OR=3.04; 95%CI=1.24-7.43; P=0.01), 9 (OR=2.72; 95%CI=1.26-5.88; P=0.009), and 14 (OR=2.10; 95%CI=1.01-4.35; P<0.05) years old were all significantly high.

In Table 5-B, the study reported by Ross et al. showed that there were significant differences for children who habituated between the GM/IVH group and the full-term control group (OR=0.1; 95%CI=0.02-0.39; P=0.0003). This means the GM/IVH group (N=14) had significantly smaller number of children who habituated than the full-term control group (N=27).

The significantly different results in Tables 5-A and 5-B were consistent with the above comparison of frequencies. On the other hand, studies reported by Papile et al. (1983), Lowe et al. (1990) and Sherlock et al. (2005) showed no significant difference in the abnormal outcomes between the GM/IVH groups and the normal ultrasound groups.

VIII. Averages and Standard Deviations of Developmental Tests

Tables 4-A and 4-B listed the means and standard deviations (SD) of various neurodevelopmental or cognitive tests. In Table 3-A, Sherlock et al. (2005) found that the normal ultrasound group had significantly lower scores than the GM/IVH group in Wide Range Achievements Test (WRAT3). The p-values for reading, spelling, and arithmetic were 0.004, 0.048, and 0.026, respectively. The results derived from WRAT3, which were reported by Sherlock et al., do not conform to common anticipation. The rest of the studies

in Tables 4-A showed that there were no significant differences between the GM/IVH groups and the normal ultrasound groups.

Table 3-B lists the means and standard deviations (SD) of various cognitive tests in studies with a full-term control group. Ross et al. (1992) reported that the GM/IVH group had significantly lower scores on Bayley Mental Developmental Index than both the premature normal ultrasound group and the full-term control group. In addition, Ross reported that there were significant differences between not only the GM/IVH group and the premature normal ultrasound group ($P<0.03$) but also the GM/IVH group and the full-term control group ($P<0.05$) on the habituation test. Ross et al. (1996) found that the GM/IVH group had significantly lower scores than both the premature normal ultrasound group ($P<0.01$) and the full-term control group ($P<0.01$) for the invisible displacement trials score of invisible displacement tasks. On the other hand, they reported that the GM/IVH group had significantly lower scores only than the full-term control group for the systematic search trials score of invisible displacement tasks ($P<0.05$). For the object discrimination reversal task in the same study, the GM/IVH group had significantly lower scores than both the premature normal ultrasound group ($P<0.05$) and the full-term control group ($P<0.05$). The study published by Lowe et al. (1990) did not show any significance for all of the cognitive tests between the GM/IVH group and the premature normal ultrasound group, as well as the GM/IVH group and the full-term control group.

IX. Overall Evaluation and Interpretation

In order to overcome the limitation due to lack of specific outcomes, I aggregated various outcome variables that appeared in the selected studies into three categories: neurodevelopment, general cognition, and special cognition. Because of the high heterogeneity among the included studies, it is infeasible to utilize meta-analysis. Therefore I calculated the weighted odds ratios (Tables 7-A and 7-B) to find a trend of the effect of GM/IVH on children's development and cognition as follows. Firstly, a sum-of-product is generated by adding each odds ratio of the outcome variable in a specific category multiplied by its sample size. Secondly, the sum-of-product is divided by the total sample sizes of the outcome variables, which produces the weighted odds ratio. In addition, the median and the range of the ORs were given, because the median of ORs can reduce the effect of extreme values of ORs, and the range of the ORs can be used to estimate the difference among the ORs. Furthermore, the weighted means of developmental tests were provided through similar methods (Tables 7-C, D, E, and F).

Eleven outcome variables (e.g. handicap, disability, and cerebral palsy etc.) that belonged to neurodevelopment were aggregated (Table 7-A). The median of ORs is 2.1 and the range of ORs is 5.96. In addition, weighted odds ratio is 2.53. I observed that the weighted OR slightly differs from the median of ORs. This means the weighted OR is slightly influenced by the extreme value of ORs. That the weighted odds ratio is 2.53 means children with GM/IVH have a risk of abnormal neurodevelopment 2.53 times higher than those without.

To evaluate the effect of GM/IVH on children's general cognitive outcomes, Table 7-B shows that the median of ORs is 1.4 and the range of ORs is 4.23. In addition, the weighted odds ratio is 2.49. I observe that the weighted OR differs from the median ORs, suggesting that the weighted OR is influenced by the extreme value of ORs. Based on the weighted OR, children with GM/IVH have a risk of abnormal general cognition 2.49 times higher than those without.

Tables 7-C, D, E, and F present un-weighted means, weighted means, and differences in weighted means. With the differences in weighted means, the effect size of GM/IVH can be inferred. In Table 7-C, the weighted means of the GM/IVH group and the normal ultrasound group in children's neurodevelopment are 97.4 and 100.7 respectively. Table 7-D shows that the effect size of GM/IVH on children's general cognition and the difference in means is between 1.2 and 2.0. In addition, Table 7-F shows that the effect size of GM/IVH on children's reasoning and memory and its difference in means is 1.98. The results show that it is possible that GM/IVH has some unfavorable influence on children's neurodevelopment, general cognition, and reasoning and memory. However, Table 7-E provides an opposite result that children with GM/IVH may obtain higher scores than those without. The difference in weighted means of beneficial effect size of GM/IVH is 2.42.

Chapter Four
Discussion

I. Evaluation of Causal Criteria

By applying Hill's causal criteria to evaluate the selected studies, I found that five of the criteria (i.e. specificity, biological plausibility, biological gradient, experiment, and analogy) were not established and one (i.e. temporality) was established in all studies. The other three criteria, strength, consistency and coherence, were thus adopted to evaluate the causal nature of the association between GM/IVH and developmental outcomes in the selected studies. Table 8 shows the evaluation of causal criteria for each included study.

Strength— "Strength" is established by eight of the selected studies [1] [12] [78]-[80] [82] [83] [85], and it means that children with GM/IVH have higher risk than those with normal ultrasound for the abnormal outcomes. The results abstracted from the studies of van de Bor et al in 1988, 1993 and 2004 reveal that the odds ratio from handicap in children with GM/IVH is twofold (OR=2.1) to threefold (OR=2.7) the odds ratio in children with normal ultrasound (Table 8-A). In Kitchen's study, the odds ratio from cerebral palsy in children with GM/IVH is over six times (OR= 6.89) the odds ratio in children with normal ultrasound (Table 8-A), while that from disability is fourfold (OR= 4.04). From the comparison between the GM/IVH group and the premature normal ultrasound group (Table 8-A), I found that the odds ratios from mental retardation (OR=5.0) and abnormal neurodevelopment (OR=4.9) are both fivefold the normal group, as reported in studies of

Whitaker [12] and Hanigan et al [78]. In 1992 and 1996, Ross et al. conducted studies to evaluate special cognition (i.e. habituation test, AB task, and invisible displacement task) for infants, and they found that the "strength" existed when the GM/IVH group was compared with the normal ultrasound group (Tables 3-B and 4-B).

Based on Cohen's scale [94] and Hopkins's scale [96], an OR that ranges from 1.5 to 3.5 is regarded as having a "small" effect size, 3.5 to 9.0 has a "moderate" effect size, and 9.0 to 32 has a "large" effect size. Therefore, the included studies, whose strength criterion is established, have the effect size between small and moderate levels.

Strength was established among half of the 16 included studies, and it may support the causation between GM/IVH and developmental outcomes. Because these selected studies lack sufficient evaluations for specific outcomes, their various outcomes were aggregated into three categories: neurodevelopment, general cognition and special cognition. By evaluating the pooled outcome variables in each category, the causal criterion of strength is moderately established, and it is necessary to obtain more evidence from related studies to further strengthen causation.

Consistency— Table 8 shows that children's disability reported by Kitchen et al. [85] was consistent with that reported by van de Bor et al. [83] (Table 8-A). In addition, studies performed by van de Bor et al. (1988; 2004) also show consistent results (Tables 3-A and 5). Furthermore, the results of children's neurodevelopment reported in the study of Hanigan et al. [78] agreed with that of van de Bor et al. [82] (Tables 3-A and 5). This observation

shows that the outcomes of the GM/IVH groups are significantly different from those of the normal ultrasound groups. Due to the lack of sufficient evaluations for specific outcomes, the causal criterion of consistency was built in the selected studies only to a certain extent.

Coherence— While most included studies revealed that either GM/IVH is not associated with children's developmental outcomes or the GM/IVH group has a higher risk for developmental outcomes than the normal ultrasound group (Table 8), Sherlock et al. (2005) found that the normal ultrasound group had significantly lower scores than the GM/IVH group in Wide Range Achievements Test (WRAT3) (Table 3-A). This result did not conform to the common anticipation. However, fifteen of the sixteen selected studies produced results which conformed to the general epidemiological and medical knowledge. Therefore, the causation of coherence was strongly established among these fifteen studies.

Overall, the studies conducted by Kitchen [85], van de Bor [82] [83] and Hanigan [78] et al. included the most evidence to support the causation because "strength", "consistency" and "coherence" were all established (Table 8). Therefore, the causation between GM/IVH and developmental outcomes in these four studies is stronger and more credible than the other studies.

II. Limitations

That it is difficult to obtain a larger number of publications for the topic of this thesis is the first limitation. Previous researchers did not give enough attention to the effect of GM/IVH on children's development and cognition. Therefore, the lack of abundant

literature undermines the establishment of stronger inference for the association between GM/IVH and neurodevelopment, as well as the association between GM/IVH and cognition.

Secondly, publication bias, one type of selection bias, is a major limitation for systematic reviews [90]. This is because studies with statistically significant results are more likely to be published than those without significant results [90]. This review included relevant published studies but could not include unpublished studies; therefore, it is difficult to minimize the publication bias. In addition, language bias, another type of selection bias, is manifested because all included studies in this research were published in English [91]. Selection bias, including publication bias and language bias, may threat the internal validity of this research.

Neurodevelopmental and cognitive outcomes, which were investigated in this research, exhibit various and comprehensive conditions. Based on previous studies, GM/IVH did not have significant influences on any typical outcome. As a result, most researchers selected some of the outcomes that they were interested in. Therefore, it is hard to pool a large number of publications for a typical outcome, such as intelligence deficiency or cerebral palsy.

A similar limitation for this review is that it is not easy to collect studies approaching long-term developments such as school performances or learning abilities. In this review, only one selected study assessed children at the age of 14. Therefore, evidence for the need of special education is very limited because of insufficient relevant studies.

The heterogeneity among the selected studies is a generic problem for this review. The differences in age of assessment, follow-up rate, assessment rate, and various developmental outcomes are the main causes of heterogeneity. In addition, due to various measurements used in the selected studies, this review lacks efficient and practicable methods to evaluate the validity and reliability of their outcomes. Similarly, various methods of statistic analyses in the selected studies lead to the heterogeneity that threatened the comparability. Because of heterogeneity, I tend to use descriptive statistic methods to analyze the results of the included studies in this thesis.

III. Strengths

This research is the first systematic review to study the effects of GM/IVH on neurodevelopmental and cognitive outcomes. It is not only a beginning for studying related issues but also a valuable guide for further investigations. With this systematic review, the neglected links between GM/IVH and children's development were explored. Consequently, more reliable evidence of association was established in this thesis.

Because this is a systematic review, it provides an improved reflection of reality [92]. Applying explicit scientific principles such as quantitative methods, selection criteria, causal criteria and criteria of study quality, the random and systematic errors of bias is likely to be reduced [93]. Therefore, this thesis not only avoids the disadvantages of traditional reviews but also provides more accurate inferences. Several high quality studies were pooled in this thesis, and highly credible interpretations were presented.

In order to overcome the limitation that no specific outcome was evaluated by sufficient studies, I tried to pool the various outcomes into larger categories. With this technique, a trend was found and possible inferences were derived.

IV. Conclusions

After evaluating the selected studies with quantitative methods, causal criteria and criteria of study quality, I summarize the results as follows.

Firstly, GM/IVH does seem to slightly impair children's special cognitive outcomes, especially special mental performances, such as infants' attention for an object, infants' development of memory for the location of an object, and children's reasoning and memory.

Secondly, general cognitive tests (e.g. the Full-Scale IQ from the Wechsler Intelligence Scales for Children- Third Edition and the Wechsler Preschool and Primary Scales of Intelligence) show that there is no significant difference between children who were with GM/IVH and those without it in global cognitive performances. However, GM/IVH seems to have certain effects on children's general cognition. It still needs more evidence to provide stronger support.

Thirdly, GM/IVH in preterm infants leads to an increased risk of adversely affected neurodevelopmental performance, including CP, disability, and handicap. In addition, the effect of GM/IVH on school performances of children or adolescents is unfavorable. Finally, a conclusion is made that the older the age of assessment, the better the evaluation.

V. Further Discussion

I observed that the rates of handicap and/or abnormal neurodevelopment differ between the GM/IVH group and the normal ultrasound group, but there is usually no difference in general cognitive scores, such as the mean IQ. In order to explain this phenomenon, I list some possible reasons as follows:

12 studies examined both outcomes except the studies of Papile et al. (1983), van de Bor et al. (1988; 1993), and Ment et al. (1985). Therefore, this phenomenon is not due to that the outcomes were examined by different studies.

Ross et al., in 1992 and 1996, conducted the studies that excluded "infants with congenital malformations, moderate to severe neurosensory deficits, or possible exposure to substance abuse *in utero*". They found the effect of GM/IVH on children's cognition is significantly unfavorable. The study of Pinto-Martin et al. (1999) excluded children with major motor or cognitive disability, and they did not find any cognitive outcomes that differ between children with GM/IVH and those without. Other studies that examined both outcomes did not provide the exclusion criteria for enrolling children. Therefore, it is hard to know whether the following two explanations caused this phenomenon: (1) children with developmental abnormalities were excluded in cognitive scores, and (2) the number of children with abnormal development was too small to affect the overall cognitive mean.

A possible supposition is that GM/IVH may influence various neurodevelopmental functions. In addition, the assessment rate may affect the reliability of a study: it is possible that the loss-of-follow-up children had poor cognitive performance or abnormal

neurodevelopment. To find the correct answer to this question, more relevant researches need to be conducted.

TABLES

Table 1-A (1 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies without a Full-Term Control Group

Author/ Year of Publication	Study Location	Number of Being Admitted Cohort	Number of who survived and can be contacted.	Admitted Criterion: GA (wks)
Papile/ 1983	Albuquerque, N.M., U.S.A.	260	232	N/A
Ment/ 1985	New Haven, C.T., U.S.A.	218	164	N/A
Kitchen/ 1990	Melbourne, Parkville, Australia.	227	154	N/A
Sostek/ 1987	Washington, D.C., U.S.A.	N/A	113	$\leq$ 34
van de Bor/ 1988	Netherlands	484	294	< 32
van de Bor/ 1993	Netherlands	484	304	< 32
van de Bor/ 2004	Netherlands	484	304	< 32
Whitaker/ 1996	New Jersey, U.S.A.	1105	898	N/A
Whitaker/ 1997	New Jersey, U.S.A.	1105	898	N/A
Pinto-Martin/ 1999	New Jersey, U.S.A.	1105	Age 2: 887 Age 6: 873 Age 9: 868	N/A
Hanigan/ 1991	U.S.A.	459	336	N/A

GA: Gestational Age; BW: Birth Weight; SD: Standard Deviation; wks: weeks; g: grams; N/A: Not Available; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; †: Significant; ‡: Slightly significant.

Table 1-A (2 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies without a Full-Term Control Group

Author/ Year	Admitted Criterion: BW (g)	GA of Groups: Mean±SD (wks)	BW of Groups: Mean±SD (g)
Papile/ 1983	< 1501	*Normal: 30.2 GM/ IVH: 29.3	*Normal: 1180 GM/ IVH: 1086
Ment/ 1985	≦ 1250	*Normal: 29.7 ± 2.3 All IVH: 28.5 ± 1.7 †	*Normal: 1020 ± 165 All IVH: 1007 ± 154
Kitchen/ 1990	500 to 1500	All preterm: 29.3 ± 1.9	All preterm: 1179 ± 214
Sostek/ 1987	< 1750	*Normal: 29.9 ± 2.14 GM/ IVH: 29.2 ± 2.21 ‡	*Normal: 1262 ± 291 GM/ IVH: 1193 ± 306
van de Bor/ 1988	< 1500	*Normal: 29.7 ± 1.5 GM/IVH: 29.2 ± 1.5 †	*Normal: 1298 ± 319 GM/IVH: 1261 ± 278
van de Bor/ 1993	< 1500	*Normal: 29.7 ± 1.5 GM/IVH: 29.2 ± 1.5 †	*Normal: 1298 ± 319 GM/IVH: 1261 ± 278
van de Bor/ 2004	< 1500	*Normal: 29.4 ± 1.5 GM/IVH: 28.8 ± 1.5 †	*Normal: 1297 ± 329 GM/IVH: 1258 ± 291
Whitaker/ 1996	501 to 2000	*Normal: 32.1 ± 3.0 GM/IVH: 30.1 ± 3.0 †	*Normal: 1529.5 ± 342.9 GM/IVH: 1339.2 ± 367.0 †
Whitaker/ 1997	501 to 2000	*Normal: 32.1 ± 3.0 GM/IVH: 30.1 ± 3.0 †	*Normal: 1529.5 ± 342.9 GM/IVH: 1339.2 ± 367.0 †
Pinto-Martin/ 1999	501 to 2000	*Normal: 32.1 ± 3.0 GM/IVH: 30.1 ± 3.0 †	*Normal: 1529.5 ± 342.9 GM/IVH: 1339.2 ± 367.0 †
Hanigan/ 1991	< 1500	*Normal: 30.5 ± 0.17 GM/IVH: 29.2 ± 0.34 †	*Normal: 1208.95 ± 16.1 GM/IVH: 1150.72 ± 34.8 †

GA: Gestational Age; BW: Birth Weight; SD: Standard Deviation; wks: weeks; g: grams; N/A: Not Available; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; †: Significant; ‡: Slightly significant.

Table 1-B (1 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies with a Full-Term Control Group

Author/ Year	Study Location	Number of Being Admitted Cohort	Number of who survived and can be contacted.	Admitted Criterion: GA (wks)	Admitted Criterion: BW (g)
Lowe/ 1990	Albuquerque, N.M., U.S.A.	FT Controls: N/A Preterm: 260	FT Controls:N/A Preterm:198	FT Controls:N/A Preterm:N/A	FT Cont: N/A Preterm:<1501
Levene/ 1992	U.K.	FT Controls: N/A Preterm: 200	FT Controls:N/A Preterm:155	FT Controls:N/A Preterm:N/A	FT Cont: N/A Preterm:<1501
Sherlock/ 2005	Victoria, Australia.	FT Controls: 265 Preterm: 568	FT Controls: 262 Preterm: 298	FT Controls:N/A Preterm:<28	FT Cont:>2499 Preterm:<1000
Ross/ 1992	New York, U.S.A.	FT Controls: 30 Preterm: 60	FT Controls: 30 Preterm: 60	FT Controls:N/A Preterm:28~32	FT Cont: N/A Preterm: N/A
Ross/ 1996	New York, U.S.A.	FT Controls: 30 Preterm: 60	FT Controls: 30 Preterm: 60	FT Controls:N/A Preterm:28~32	FT Cont: N/A Preterm:N/A

GA: Gestational Age; BW: Birth Weight; wks: weeks; g: grams; SD: Standard Deviation; N/A: Not Available; SES: Socio-Economic Status; FT: Full Term; Cont: Controls; *Normal: premature normal-ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage.

Table 1-B (2 of 2): The Numbers, Gestational Ages, and Birth Weights of Baseline Cohorts in Studies with a Full-Term Control Group

Author/ Year	GA of Groups: Mean±SD (wks)	BW of Groups: Mean±SD (g)	Match
Lowe/ 1990	FT Controls: N/A *Normal: 30.4 GM/ IVH: 30.5	FT Controls: N/A *Normal: 1157 GM/ IVH: 1137	Race; age; gender; SES.
Levene/ 1992	FT Controls: N/A *Normal: N/A GM/ IVH: N/A	FT Controls: N/A *Normal: N/A GM/ IVH: N/A	Age; class; school.
Sherlock/ 2005	FT Controls: 39.3 ±1.4 All preterm: 26.7 ± 2	FT Controls: 3407 ± 443 All preterm: 883 ± 162	Gender; the mother's country of birth; health insurance status.
Ross/ 1992	FT Controls: 39.4 ±0.9 *Normal: 30.5 ± 1.5 GM/ IVH: 30.2 ± 1.4	FT Controls: 3875.2 ± 321 *Normal: 1494.7 ± 256 GM/ IVH: 1431.1 ± 226	GA; birthweight (within 100 g); days of assisted ventilation (within 1 week); SES; gender; race.
Ross/ 1996	FT Controls: 39.4 ±0.9 *Normal: 30.5 ± 1.5 GM/ IVH: 30.2 ± 1.4	FT Controls: 3875.8 ± 285 *Normal: 1483.6 ± 266 GM/ IVH: 1432.8 ± 227	GA; birthweight (within 100 g); days of assisted ventilation (within 1 week); SES; gender; race.

GA: Gestational Age; BW: Birth Weight; wks: weeks; g: grams; SD: Standard Deviation; N/A: Not Available; SES: Socio-Economic Status; FT: Full Term; Cont: Controls; *Normal: premature normal-ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage.

Table 2-A (1 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies without a Full-Term Control Group

Author/ Year	Age of Assessment	Follow-up Rate (%)	Assessment Rate (%)
Papile/ 1983	12 or 24 months	198/ 232= 85.34%	197/ 232= 84.91%
Ment/ 1985	12, 18, and 30 months	142/ 164= 86.59%	142/ 164= 86.59%
Kitchen/ 1990	5 years	139/ 154= 90.26%	135/ 154= 87.66%
Sostek/ 1987	Mean age at 12.6 months and 21.5 months	At 12.6 months 89/ 113= 78.76% At 21.5 months 86/ 113= 76.11%	At 12.6 months 89/ 113= 78.76% At 21.5 months 86/ 113= 76.11%
van de Bor/ 1988	2 years	294/ 294= 100.00%	294/ 294= 100.00%
van de Bor/ 1993	5 years	304/ 304= 100.00%	301/ 304= 99.01%
van de Bor/ 2004	14 years	278/ 304= 91.45%	278/ 304= 91.45%
Whitaker/ 1996	6 years	685/ 898= 76.28%	597/ 898= 66.48%
Whitaker/ 1997	6 years	685/ 898= 76.28%	564/ 898= 62.81%
Pinto-Martin/ 1999	 2 years 6 years 9 years 2 years 6 years 9 years	Motor Ability 777/ 887= 87.59% 685/ 873= 78.47% 658/ 868= 75.81% General Cognition 777/ 887= 87.59% 685/ 873= 78.47% 658/ 868= 75.81%	Motor Ability 462/ 887= 52.09% 538/ 873= 61.63% 487/ 868= 56.11% General Cognition 611/ 887= 68.88% 538/ 873= 61.63% 488/ 868= 56.22%
Hanigan/ 1991	18 months and 3 years	219/ 336= 65.18%	216/ 336= 64.29%

N: Number; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; VE/PL: Ventricular Enlargement and/or Parenchymal Lesion; N/A: Not Available.

Table 2-A (2 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies without a Full-Term Control Group

Author/ Year	Number of Assessment			
	Total N	*Normal	GM/ IVH	VE/ PL
Papile/ 1983	197	115	51	31
Ment/ 1985	142	94	62	8
Kitchen/ 1990	135	96	33	6
Sostek/ 1987	89	36	23	30
	86	36	20	30
van de Bor/ 1988	294	225	52	17
van de Bor/ 1993	301	234	50	17
van de Bor/ 2004	278	216	45	17
Whitaker/ 1996	597	468	83	46
Whitaker/ 1997	564	454	78	32
Pinto-Martin/ 1999	462	382	61	19
	538	449	70	19
	487	402	65	20
	611	504	83	24
	538	449	70	19
	488	403	65	20
Hanigan/ 1991	216	153	47	16

N: Number; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; VE/PL: Ventricular Enlargement and/or Parenchymal Lesion; N/A: Not Available.

Table 2-B (1 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies with a Full-Term Control Group

Author/ Year	Age of Assessment	Follow-up Rate (%)	Assessment Rate (%)
Lowe/ 1990	From 5 to 6 years	FT Controls: N/A Preterm: 38/ 198= 19.19%	FT Controls: N/A Preterm: 38/ 198= 19.19%
Levene/ 1992	5 years	FT Controls: N/A Preterm: 152/ 155= 98.06%	FT Controls: N/A Preterm: 140/ 155= 90.32%
Sherlock/ 2005	8 years	FT Controls: 221/262=84.35% Preterm: 275/ 298= 92.28%	FT Controls: 221/262= 84.35% Preterm: 270/ 298= 90.60%
Ross/ 1992	10 months	FT Controls: 30/ 30= 100% Preterm: 60/ 60= 100%	FT Controls: 30/ 30= 100% Preterm: 60/ 60= 100%
Ross/ 1996	2 years	FT Controls: 27/ 30= 90% Preterm: 55/ 60= 91.67%	FT Controls: 27/ 30= 90% Preterm: 55/ 60= 91.67%

FT: Full Term; N/A: Not Available; N: Number; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; VE/PL: Ventricular Enlargement and/or Parenchymal Lesion; †: the sum of prolonged flare (PF) group and GM/IVH with PF group.

Table 2-B (2 of 2): Follow-up Rates, Assessment Rates, and Number of Assessment in Studies with a Full-Term Control Group

Author/ Year	Number of Assessment				
	Total N	FT Controls	*Normal	GM/ IVH	VE/ PL
Lowe/ 1990	60	22	27	11	N/A
Levene/ 1992	284	144	64	54	22†
Sherlock/ 2005	491	221	180	72	18
Ross/ 1992	90	30	30	30	N/A
Ross/ 1996	82	27	28	27	N/A

FT: Full Term; N/A: Not Available; N: Number; *Normal: premature normal ultrasound group; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; VE/PL: Ventricular Enlargement and/or Parenchymal Lesion; †: the sum of prolonged flare (PF) group and GM/IVH with PF group.

Table 3-A (1 of 2): Study Quality Evaluation

Author/ Year	Follow-up rate ≧ 90%	Assessment rate ≧ 85%	Ms of GA are NS.	Ms of BW are NS.
Papile/ 1983	0	1	1	1
Ment/ 1985	0	1	0	1
Kitchen/ 1990	1	1	1	1
Sostek/ 1987	0	0	0	1
van de Bor/ 1988	1	1	0	1
van de Bor/ 1993	1	1	0	1
van de Bor/ 2004	1	1	0	1
Whitaker/ 1996	0	0	0	0
Whitaker/ 1997	0	0	0	0
Pinto-Martin/ 1999	0	0	0	0
Hanigan/ 1991	0	0	0	0
Lowe/ 1990	0	0	1	1
Levene/ 1992	1	1	1	1
Sherlock/ 2005	1	1	1	1
Ross/ 1992	1	1	1	1
Ross/ 1996	1	1	1	1

1: Its answer for a criterion is "yes"; 0: Its answer for a criterion is "no"; Ms: Means; GA: Gestational Age; NS: Not Significant; US: Ultrasound; FT: Full-Term.

Table 3-A (2 of 2): Study Quality Evaluation

Author/ Year	Utilizing US scan	Has well matched FT controls	Sample size of each study group (case) ≧30
Papile/ 1983	1	0	1
Ment/ 1985	1	0	1
Kitchen/ 1990	1	0	1
Sostek/ 1987	1	0	0
van de Bor/ 1988	1	0	1
van de Bor/ 1993	1	0	1
van de Bor/ 2004	1	0	1
Whitaker/ 1996	1	0	1
Whitaker/ 1997	1	0	1
Pinto-Martin/ 1999	1	0	1
Hanigan/ 1991	1	0	1
Lowe/ 1990	0	1	0
Levene/ 1992	1	1	1
Sherlock/ 2005	1	1	1
Ross/ 1992	1	1	1
Ross/ 1996	1	1	0

1: Its answer for a criterion is "yes"; 0: Its answer for a criterion is "no"; Ms: Means; GA: Gestational Age; NS: Not Significant; US: Ultrasound; FT: Full-Term.

Table 3-B (1 of 2): Study Quality Evaluation

Author/ Year	Has careful or standard measurements of outcomes	Age of Assessment	Key confounders are mentioned and adjusted.
Papile/ 1983	1	1	0
Ment/ 1985	1	2	0
Kitchen/ 1990	1	3	0
Sostek/ 1987	1	1	0
van de Bor/ 1988	1	2	1
van de Bor/ 1993	1	3	1
van de Bor/ 2004	1	5	1
Whitaker/ 1996	1	4	1
Whitaker/ 1997	1	4	1
Pinto-Martin/ 1999	1	4	1
Hanigan/ 1991	1	2	0
Lowe/ 1990	1	4	0
Levene/ 1992	1	3	0
Sherlock/ 2005	1	4	0
Ross/ 1992	1	1	0
Ross/ 1996	1	2	0

1: Its answer for a criterion is "yes"; 0: Its answer for a criterion is "no"; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage. Age of assessment: (1) $\geq$10 months & $\leq$1.5 years: 1 point; (2) >1.5 years & $\leq$3 years: 2 points; (3) >3 years & $\leq$5 years: 3 points; (4) >5 years & $\leq$10 years: 4 points; (5) >10 years & $\leq$18 years: 5 points.

Table 3-B (2 of 2): Study Quality Evaluation

Author/ Year	The other risk factors are mentioned and adjusted	Has a clear conclusion about GM/IVH	Total scores of study quality
Papile/ 1983	0	1	8
Ment/ 1985	0	1	8
Kitchen/ 1990	0	0	10
Sostek/ 1987	0	0	4
van de Bor/ 1988	1	1	11
van de Bor/ 1993	1	1	12
van de Bor/ 2004	1	1	14
Whitaker/ 1996	1	1	10
Whitaker/ 1997	1	1	10
Pinto-Martin/ 1999	1	0	9
Hanigan/ 1991	0	0	5
Lowe/ 1990	1	1	10
Levene/ 1992	0	1	12
Sherlock/ 2005	0	1	13
Ross/ 1992	1	1	11
Ross/ 1996	1	1	11

1: Its answer for a criterion is "yes"; 0: Its answer for a criterion is "no"; GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage.

Table 4 (1 of 2): Ranking of Importance for Selected Studies

Author/ Year	Ranking of Importance	Total scores of study quality
van de Bor/ 2004	1	14
van de Bor/ 1993	2	12
van de Bor/ 1988	3	11
Sherlock/ 2005	4	13
Levene/ 1992	4	12
Ross/ 1992	4	11
Ross/ 1996	4	11
Kitchen/ 1990	4	10
Whitaker/ 1996	5	10
Whitaker/ 1997	6	10
Lowe/ 1990	6	10
Pinto-Martin/ 1999	7	9
Papile/ 1983	8	8
Ment/ 1985	8	8
Hanigan/ 1991	9	5
Sostek/ 1987	10	4

Ranking of importance from 1 to 10 (1 means the most important study; 10 means the last important study). Plus mark (+) means the criterion is established. Minus mark (−) means the criterion is not established.

Table 4 (2 of 2): Ranking of Importance for Selected Studies

Author/ Year	Ranking of Importance	Causal Criteria		
		Strength	Consistency	Coherence
van de Bor/ 2004	1	+	−	+
van de Bor/ 1993	2	+	+	+
van de Bor/ 1988	3	+	+	+
Sherlock/ 2005	4	−	−	−
Levene/ 1992	4	−	−	+
Ross/ 1992	4	+	−	+
Ross/ 1996	4	+	−	+
Kitchen/ 1990	4	+	+	+
Whitaker/ 1996	5	+	−	+
Whitaker/ 1997	6	−	−	+
Lowe/ 1990	6	−	−	+
Pinto-Martin/ 1999	7	−	−	+
Papile/ 1983	8	−	−	+
Ment/ 1985	8	−	−	+
Hanigan/ 1991	9	+	+	+
Sostek/ 1987	10	−	−	+

Ranking of importance from 1 to 10 (1 means the most important study; 10 means the last important study). Plus mark (+) means the criterion is established. Minus mark (−) means the criterion is not established.

Table 5-A (1 of 2): Frequencies and Odds Ratios of Abnormal Outcomes in Studies without a Full-Term Control Group

Author/ Year	Abnormal Outcomes	GM/ IVH (%)
Papile/ 1983	**Bayley Scales of Infant Development**	
	Abnormal outcome	5/51= 9.80%
	Neurodevelopmental Examination	
	Minor handicap	21/51= 41.18%
	Major handicap	5/51= 9.8%
Kitchen/ 1990	CP	6/33= 18.18%
	Disability	6/33= 18.18%
Sostek/ 1987	At 12.6 months	
	Mental Delay	8.6%
	Motor Delay	30.5%
van de Bor/ 1988	‡Neurodevelopmental outcome (Handicap)	19/52= 36.5%
van de Bor/ 1993	‡Disability (including Handicap)	21/50=42%
	‡Total Handicap	13/50= 26%
van de Bor/ 2004	Disability or Handicaps at 5 Years Overall	
	None	29/45= 64.44%
	Disability	16/45= 35.56%
	Handicap	11/45= 24.44%
	School Performance at 5 Years	
	Mainstream	31/40= 77.5%
	Special Education	9/40= 22.5%
	School Performance at 9 Years	
	Normal	22/40= 55.0%
	Slow Learner	5/40= 12.5%
	Special Education	13/40= 32.5%
	School Performance at 14 Years	
	Normal	25/45= 55.6%
	Slow Learner	10/45= 22.2%
	‡Special Education	10/45= 22.2%
Whitaker/ 1996	‡Borderline Intelligence	5/83= 6.0%
	‡MR	5/83= 6.0%
Whitaker/ 1997	‡Any disorder in DISC 2.1 P	17/78= 21.8%
Hanigan/ 1991	Abnormal Neurodevelopment	12/47= 25.53%

Lowe/ 1990	Abnormality on McCSCA	4/11=36.36%
	Abnormality on TERA	5/11=45.45%
	Abnormality on DT-VMI	2/11=18.18%
Sherlock/ 2005	Abnormal Movement ABC	17/64= 26.56%
	CP	9/72= 12.50%
	Major neurosensory disability	10/72= 13.89%

GM/ IVH: Germinal Matrix and/or Intraventricular Hemorrhage; *Normal: premature normal ultrasound group; CP: Cerebral Palsy; MR: Mental Retardation; DISC 2.1 P: the Diagnostic Interview Schedule for Children-Parent version 2.1P; McCSCA: the McCarthy Scales of Children's Ability; TERA: the Test of Early Reading Ability; DT-VMI: Developmental Test of Visual-Motor Integration; N/A: Not Available; Movement ABC: the Movement Assessment Battery for Children; OR: Odds Ratio; CI: Confidence Interval; NS: Not Significant; †: significant; The frame mark (☐) means it contributed to the cell of two-by-two table has a expected count less than 5; ‡: the odds ratio, 95% Confidence Interval, and p-value were provided by the publication itself.

Table 5-A (2 of 2): Frequencies and Odds Ratios of Abnormal Outcomes in Studies without a Full-Term Control Group

Author/ Year	*Normal (%)	OR; (95% CI); P-value		
Papile/ 1983				
	11/115= 9.57%	OR=1.03	(0.34-3.13)	NS
	46/115= 40%	OR=1.05	(0.54-2.05)	NS
	12/115= 10.43%	OR=0.93	(0.31-2.80)	NS
Kitchen/ 1990	3/96= 3.13 %	OR=6.89	(1.62-29.39)	P=0.003†
	5/96= 5.21 %	OR=4.04	(1.15-14.29)	P=0.02†
Sostek/ 1987				
	16.7%	N/A		
	22.2%	N/A		
van de Bor/ 1988	40/225= 17.8%	OR=2.10	(1.30-3.30)	P<0.01†
van de Bor/ 1993	57/234= 24%	OR=2.20	(1.12-4.40)	P<0.05†
	31/234= 13%	OR=1.90	(0.81-4.34)	NS
van de Bor/ 2004				
	164/216= 75.92%			
	49/216= 22.68%	OR=1.88	(0.95-3.74)	NS
	23/216= 10.65%	OR=2.70	(1.21-6.08)	P=0.01†
	178/195= 91.3%			
	17/195= 8.7%	OR=3.04	(1.24-7.43)	P=0.01†
	107/193= 55.4%			
	57/193= 29.5%	OR=0.34	(0.13-0.91)	P=0.02†
	29/193= 15.0%	OR=2.72	(1.26-5.88)	P=0.009†
	97/216= 44.9%			
	93/216= 43.1%	OR=0.38	(0.18-0.80)	P=0.009†
	26/216= 12.0%	OR=2.10	(1.01-4.35)	P<0.05†
Whitaker/ 1996	26/468= 5.6%	OR=1.10	(0.40-3.10)	NS
	6/468= 1.3%	OR=5.00	(1.50-16.80)	P<0.01†
Whitaker/ 1997	93/454= 20.3%	OR=1.40	(0.70-2.70)	NS

Hanigan/ 1991	10/153= 6.54%	OR=4.90	(1.96-12.27)	P=0.0003†
Lowe/ 1990	5/27=18.52%	OR=2.51	(0.52-12.04)	NS
	14/27=51.85%	OR=0.77	(0.19-3.16)	NS
	8/27=29.63%	OR=0.53	(0.09-3.01)	NS
Sherlock/ 2005	39/173= 22.54%	OR=1.24	(0.64-2.40)	NS
	12/180= 6.67%	OR=2.00	(0.80-4.98)	NS
	28/180= 15.56%	OR=0.88	(0.40-1.91)	NS

GM/ IVH: Germinal Matrix and/or Intraventricular Hemorrhage; *Normal: premature normal ultrasound group; CP: Cerebral Palsy; MR: Mental Retardation; DISC 2.1 P: the Diagnostic Interview Schedule for Children-Parent version 2.1P; McCSCA: the McCarthy Scales of Children's Ability; TERA: the Test of Early Reading Ability; DT-VMI: Developmental Test of Visual-Motor Integration; N/A: Not Available; Movement ABC: the Movement Assessment Battery for Children; OR: Odds Ratio; CI: Confidence Interval; NS: Not Significant; †: significant; The frame mark (☐) means it contributed to the cell of two-by-two table has a expected count less than 5; ‡: the odds ratio, 95% Confidence Interval, and p-value were provided by the publication itself.

Table 5-B (1 of 2): Frequencies and Odds Ratios of Special Cognitive Outcomes in a Study with a Full-Term Control Group

Author/ Year	Measures	GM/ IVH (%)	*Normal (%)	FT Controls (%)
Ross/ 1992	Children who habituated	14/30= 46.67%	29/30= 96.67%	27/30= 90.00%
	Response on AB task			
	(a) Failed reversal trials	22/30= 73.33%	17/30= 56.67%	5/30= 16.67%
	(b) AB error without delay	6/30= 20.00%	8/30= 26.67%	4/30= 13.33%
	(c) 1-3 seconds delay	2/30= 6.67%	4/30= 13.33%	17/30= 56.67%
	(d) ≧ 5 seconds delay	0/30= 0.00%	1/30= 3.33%	4/30= 13.33%

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; *Normal: premature normal ultrasound group; FT: Full-Term; OR: Odds Ratio; CI: Confidence Interval; †: significant; ‡: slightly significant; NS: Not Significant; The frame mark (☐) means it contributed to the cell of two-by-two table has a expected count less than 5.

Table 5-B (2 of 2): Frequencies and Odds Ratios of Special Cognitive Outcomes in a Study with a Full-Term Control Group

Author/ Year	OR; (95% CI); P-value (Compared with *Normal)			OR; (95% CI); P-value (Compared with FT Controls)		
Ross/ 1992	OR=0.03	(0.004-0.25)	P<.0001†	OR=0.10	(0.02-0.39)	P=.0003†
	OR=2.10	(0.71-6.22)	NS	OR=13.75	(3.92-48.27)	P<.0001†
	OR=0.69	(0.21-2.30)	NS	OR=1.63	(0.41-6.47)	NS
	OR=0.46	(0.08-2.75)	NS	OR=0.06	(0.01-0.27)	P<.0001†
	OR=0.32	(0.01-8.24)	NS	OR=0.10	(0.01-1.88)	P=0.04‡

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; *Normal: premature normal ultrasound group; FT: Full-Term; OR: Odds Ratio; CI: Confidence Interval; †: significant; ‡: slightly significant; NS: Not Significant; The frame mark (☐) means it contributed to the cell of two-by-two table has a expected count less than 5.

Table 6-A (1 of 2): Means and Standard Deviations of Developmental Tests in Studies without a Full-Term Control Group

Author/ Year	Measures and/or outcomes	GM/IVH Mean (SD); N	
Kitchen/ 1990	WPPSI Full Scale	107.50 (11.85)	N= 33
Ment/ 1985	Bayley 12 Mental	93.05 (15.23)	N= 49
	Bayley 18 Mental	88.93 (17.37)	N= 42
	SBIS 30 Month CA	87.02 (14.85)	N= 40
	PPVT 30 Month CA	88.58 (17.86)	N= 35
Sostek/ 1987	At 12.6 months		
	Bayley Scales of Infant Development		
	(a) Mental	105.80 (16.10)	N= 23
	(b) Motor	93.60 (22.60)	N= 23
	At 21.5 months		
	Bayley Scales of Infant Development		
	(a) Mental	105.20 (16.90)	N= 20
	(b) Motor	96.40 (20.50)	N= 20
Whitaker/ 1996	Language		
	(a) Overall (TOLD SLQ)	94.60 (18.40)	N= 73
	(b) Receptive (TOLD LIQ)	95.50 (15.20)	N= 73
	(c) Expressive (TOLD SPQ)	94.50 (19.60)	N= 73
	(d) Verbal reasoning (SB area score)	102.90 (10.50)	N= 73
	Short-term memory (SB area score)	99.20 (11.00)	N= 73
	Quantitative reasoning (SB area score)	107.20 (9.10)	N= 73
	Visual perceptual organization		
	(a) Abstract visual reasoning (SB area	101.20 (10.40)	N= 73
	score)	92.20 (10.70)	N= 73
	(b) Visual-motor integration	105.70 (20.60)	N= 71
	(c) TVPS perceptual quotient		
Pinto-Martin/	Motor Ability		
1999	Age 2	98.90 (15.80)	N= 61
	Age 6	7.00 (2.90)	N= 70
	Age 9	7.90 (2.80)	N= 65
	General Cognitive Ability		
	Age 2	105.20 (18.80)	N= 83

	Age 6	102.90 (10.40)	N= 70
	Age 9	99.40 (12.80)	N= 65
Levene/ 1992	TOMI- Overall: Median (Ranges)	3.00 (0-13)	N= 54
	WPPSI (vocabulary): Mean (SD)	19.20 (6.64)	N= 54
Sherlock/ 2005	Cognitive and educational outcomes		
	(a) For IQ Score	104.17 (15.32)	N= 72
	(b) For WRAT3		
	Reading	100.85 (14.80)	N= 69
	Spelling	96.85 (11.55)	N= 68
	Arithmetic	93.10 (12.75)	N= 68

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; SD: Standard Deviation; *Normal: premature normal ultrasound group; N: Number; WPPSI: the Wechsler Preschool and Primary Scales of Intelligence; SBIS: Stanford-Binet Intelligence Test; PPVT: Peabody Picture Vocabulary Test; CA: Corrected Age; TOLD : Test of Language Development; SLQ: speaking quotient; LIQ: listening quotient; SB: the Composite Index of the Stanford Binet; TVPS: the Test of Visual Perceptual Skills; IQ: Intelligence Quotient; SAS: Standard Age Score; ECA: Expressive Communication Age; RCA: Receptive Communication Age; TOMI: the Test of Motor Impairment; WRAT3: Wide Range Achievements Test; NS: Not Significant; †: premature normal ultrasound group had significantly lower score than GM/IVH group.

Table 6-A (2 of 2): Means and Standard Deviations of Developmental Tests in Studies without a Full-Term Control Group

Author/ Year	*Normal Mean (SD); N	Significance
Kitchen/ 1990	106.30 (14.20) N= 96	NS
Ment/ 1985	95.5 (13.1) N= 65	NS
	90.9 (16.0) N= 5	NS
	88.2 (17.2) N= 47	NS
	88.2 (17.0) N= 36	NS
Sostek/ 1987		
	105.20 (23.50) N= 36	NS
	96.40 (25.00) N= 36	NS
	103.10 (27.50) N= 36	NS
	93.70 (24.80) N= 36	NS
Whitaker/ 1996		
	95.60 (16.10) N= 429	NS
	97.20 (13.70) N= 428	NS
	94.70 (17.10) N= 432	NS
	104.30 (10.90) N= 436	NS
	101.40 (11.40) N= 436	NS
	108.80 (10.00) N= 436	NS
	102.10 (13.00) N= 436	NS
	93.10 (10.70) N= 431	NS
	106.90 (19.70) N= 421	NS
Pinto-Martin/ 1999		
	101.70 (14.80) N= 382	NS
	7.00 (3.00) N= 449	NS
	7.80 (3.00) N= 402	NS
	106.20 (19.80) N= 504	NS
	103.70 (12.10) N= 449	NS

	99.10 (14.50)	N= 403	NS
Levene/ 1992	2.25 (0-16)	N= 64	NS
	19.30 (5.82)	N= 64	NS
Sherlock/ 2005			
	104.19 (15.35)	N= 180	NS
	95.20 (15.70)	N= 172	P=0.004†
	93.60 (12.40)	N= 172	P=0.048†
	88.30 (14.30)	N= 170	P=0.026†

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; SD: Standard Deviation; *Normal: premature normal ultrasound group; N: Number; WPPSI: the Wechsler Preschool and Primary Scales of Intelligence; SBIS: Stanford-Binet Intelligence Test; PPVT: Peabody Picture Vocabulary Test; CA: Corrected Age; TOLD : Test of Language Development; SLQ: speaking quotient; LIQ: listening quotient; SB: the Composite Index of the Stanford Binet; TVPS: the Test of Visual Perceptual Skills; IQ: Intelligence Quotient; SAS: Standard Age Score; ECA: Expressive Communication Age; RCA: Receptive Communication Age; TOMI: the Test of Motor Impairment; WRAT3: Wide Range Achievements Test; NS: Not Significant; †: premature normal ultrasound group had significantly lower score than GM/IVH group.

Table 6-B (1 of 2): Means and Standard Deviations of Cognitive Tests in Studies with a Full-Term Control Group

Author/ Year	Measures
Lowe/ 1990	(a) McCSCA Mean (Range) (b) TERA Mean (Range) (c) DT-VMI Mean (Range) (d) BFL R Mean (Range) L Mean (Range)
Ross/ 1992	(a) Bayley MDI (b) Bayley PDI (c) Habituation- Trials to habituation (d) Habituation- Increase in % time looking at novel face
Ross/ 1996	Bayley MDI Bayley PDI Habituation- Trials to habituation Habituation- Increase in % time looking at novel face Invisible displacements task- ID trials score (maximum of 8.0) Invisible displacements task- SS trials score (maximum of 4.0) Object discrimination reversal task (Number of reversals)

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; SD: Standard Deviation; *Normal: premature normal ultrasound group; FT: Full-Term; McCSCA: the McCarthy Scales of Children's Ability; TERA: the Test of Early Reading Ability; DT-VMI: Developmental Test of Visual-Motor Integration; BFL: Benton Finger Localization; R: Right; L: Left; Bayley MDI: Bayley Mental Developmental Index; Bayley PDI: Bayley Psychomotor Development Index; ID: Invisible Displacement; SS: Systematic Search; NS[a]: No significant difference on the test between GM/IVH and *Normal groups; NS[b]: No significant difference on the test between GM/IVH and FT Control groups; †: there is a significant difference on the test between GM/IVH and *Normal groups; ‡:there is a significant difference on the test between GM/IVH and FT Control groups.

Table 6-B (2 of 2): Means and Standard Deviations of Cognitive Tests in Studies with a Full-Term Control Group

Author/ Year	GM/IVH Mean (SD)	*Normal Mean (SD)	FT Controls Mean (SD)	Significance
Lowe/ 1990	95(71-120)	97(75-123)	117(89-141)	NS[a]
	90.8(66-126)	82.6(55-110)	104.1(79-139)	NS[a]
	10.8(5-15)	9.4(4-15)	11.9(9-16)	NS[a]
	7.3(3-10)	7.6(5-10)	9.2(8-10)	NS[a]
	7.1(3-10)	7.0(3-10)	9.1(7-10)	NS[a]
Ross/ 1992	104.4 (18)	113.4 (11)	110.0 (12)	P< 0.05†‡
	94.3 (14)	99.7 (16)	102.8 (13)	NS[a]; NS[b]
	25.2 (6)	20.3 (6)	21.2 (7)	P<0.03†; P< 0.05‡
	6.2 (4)	6.8 (2)	6.3 (3)	NS[a]; NS[b]
Ross/ 1996	101.8 (16.2)	105.2 (19.1)	109.4 (20.0)	NS[a]; NS[b]
	98.2 (19.0)	97.1 (15.5)	106.3 (16.3)	NS[a]; NS[b]
	21.6 (5.7)	20.6 (6.4)	20.6 (6.9)	NS[a]; NS[b]
	6.4 (10.2)	12.2 (12.4)	13.4 (15.1)	NS[a]; NS[b]
	3.9 (1.7)	5.2 (1.3)	5.3 (1.6)	P<0.01†; P<0.01‡
	1.5 (1.2)	1.6 (1.2)	2.4 (1.2)	NS[a]; P<0.05‡
	1.52 (0.5)	2.0 (0.6)	2.1 (0.5)	P<0.05†; P< 0.05‡

GM/IVH: Germinal Matrix and/or Intraventricular Hemorrhage; SD: Standard Deviation; *Normal: premature normal ultrasound group; FT: Full-Term; McCSCA: the McCarthy Scales of Children's Ability; TERA: the Test of Early Reading Ability; DT-VMI: Developmental Test of Visual-Motor Integration; BFL: Benton Finger Localization; R: Right; L: Left; Bayley MDI: Bayley Mental Developmental Index; Bayley PDI: Bayley Psychomotor Development Index; ID: Invisible Displacement; SS: Systematic Search; NS[a]: No significant difference on the test between GM/IVH and *Normal groups; NS[b]: No significant difference on the test between GM/IVH and FT Control groups; †: there is a significant difference on the test between GM/IVH and *Normal groups; ‡:there is a significant difference on the test between GM/IVH and FT Control groups.

Table 7-A Weighted OR for Abnormal Neurodevelopment

Author/ Year	Measures/ Outcomes	OR
Papile/ 1983	Major handicap	0.93
Kitchen/ 1990	CP	6.89
Kitchen/ 1990	Disability	4.04
van de Bor/ 1988	Handicap	2.10
van de Bor/ 1993	Disability (including handicap)	2.20
van de Bor/ 1993	Handicap	1.90
van de Bor/ 2004	Disability	1.88
van de Bor/ 2004	Handicap	2.70
Hanigan/ 1991	Abnormal neurodevelopment	4.90
Sherlock/ 2005	Abnormal movement ABC	1.24
Sherlock/ 2005	CP	2.00
Median of ORs		2.1
Range of ORs		5.96
Un-weighted OR		2.80
Weighted OR		2.53

Table 7-B Weighted OR for General Cognition

Author/ Year	Measures/ Outcomes	OR
Whitaker/ 1996	Borderline intelligence	1.1
Whitaker/ 1996	Mental retardation	5
Whitaker/ 1997	Any disorder in DISC 2.1 P	1.4
Lowe/ 1990	Abnormality on McCSCA	2.51
Lowe/ 1990	Abnormality on TERA	0.77
Median of ORs		1.4
Range of ORs		4.23
Un-weighted OR		2.20
Weighted OR		2.49

Table 7-C Weighted Means for Neurodevelopment: Motor Abilities

Author/ Year	Measures/ Outcomes	Mean of the GM/IVH group	Sample size of the GM/IVH group	Mean of the *Normal group	Sample size of the *Normal group
Sostek/ 1987	Bayley PDI	96.4	20	93.7	36
Pinto-Martin/ 1999	Bayley PDI	98.9	61	101.7	382
Ross/ 1992	Bayley PDI	94.3	30	99.7	30
Ross/ 1996	Bayley PDI	98.2	27	97.1	28
Un-w Means		96.95		98.05	
W Means		97.40		100.70	
D in W Means				3.3	

Un-w Means: Un-weighted Means; W Means: Weighted Means; D in W Means: Difference in Weighted Means.

Table 7-D Weighted Means for General Cognition

Author/ Year	Measures/ Outcomes	Mean of the GM/IVH group	Sample size of the GM/IVH group	Mean of the *Normal group	Sample size of the *Normal group
Kitchen/ 1990	WPPSI Full Scale	107.5	33	106.3	96
Ment/ 1985	SBIS 30 Month CA	87.02	40	88.2	47
Ment/ 1985	PPVT 30 Month CA	88.58	35	88.2	36
Sostek/ 1987	Bayley MDI	105.2	20	103.1	36
Pinto-Martin/ 1999	WISC-III	99.4	65	99.1	403
Sherlock/ 2005	IQ Score	104.17	72	104.19	180
Lowe/ 1990	McCSCA	95	11	97	27
Ross/ 1992	Bayley MDI	104.4	30	113.4	30
Ross/ 1996	Bayley MDI	101.8	27	105.2	28
Un-w Means		99.23		100.52	
W Means		99.46		100.67	
D in W Means				1.21	

Un-w Means: Un-weighted Means; W Means: Weighted Means; D in W Means: Difference in Weighted Means.

Table 7-E Weighted Means for Special Cognition - Language

Author/ Year	Measures/ Outcomes	Mean of the GM/IVH group	Sample size of the GM/IVH group	Mean of the *Normal group	Sample size of the *Normal group
Whitaker/ 1996	Overall TOLD SLQ	94.6	73	95.6	429
Sherlock/ 2005	WRAT3 Reading	100.85	69	95.2	172
Sherlock/ 2005	WRAT3 Spelling	96.85	68	93.6	172
Lowe/ 1990	TERA	90.80	11	82	27
Un-w Means		95.77		91.6	
W Means		97.05		94.63	
D in W Means				2.42	

Un-w Means: Un-weighted Means; W Means: Weighted Means; D in W Means: Difference in Weighted Means.

Table 7-F Weighted Means for Special Cognition– Reasoning & Memory

Author/ Year	Measures/ Outcomes	Mean of the GM/IVH group	Sample size of the GM/IVH group	Mean of the *Normal group	Sample size of the *Normal group
Whitaker/ 1996	Short-term memory	99.2	73	101.4	436
Whitaker/ 1996	Quantitative reasoning	107.2	73	108.8	436
Whitaker/ 1996	Abstract visual reasoning	101.2	73	102.1	436
Sherlock/ 2005	WRAT3 (Arithmetic)	93.1	68	88.3	170
Un-w Means		100.18		100.15	
W Means		100.30		102.28	
D in W Means				1.98	

Un-w Means: Un-weighted Means; W Means: Weighted Means; D in W Means: Difference in Weighted Means.

Table 8 (1 of 2): Causal Criteria for Selected Studies

Author/ year	Strength	Consistency	Coherence	Specificity	Temporality
Papile/ 1983	−	−	+	−	+
Ment/ 1985	−	−	+	−	+
Kitchen/ 1990	+	+	+	−	+
Sostek/ 1987	−	−	+	−	+
van de Bor/ 1988	+	+	+	−	+
van de Bor/ 1993	+	+	+	−	+
van de Bor/ 2004	+	−	+	−	+
Whitaker/ 1996	+	−	+	−	+
Whitaker/ 1997	−	−	+	−	+
Pinto-Martin/ 1999	−	−	+	−	+
Hanigan/ 1991	+	+	+	−	+
Lowe/ 1990	−	−	+	−	+
Levene/ 1992	−	−	+	−	+
Sherlock/ 2005	−	−	−	−	+
Ross/ 1992	+	−	+	−	+
Ross/ 1996	+	−	+	−	+

Plus mark (+) means the criterion is established. Minus mark (−) means the criterion is not established.

Table 8 (2 of 2): Causal Criteria for Selected Studies

Author/ year	Biological Plausibility	Biological Gradient	Experiment	Analogy
Papile/ 1983	–	–	–	–
Ment/ 1985	–	–	–	–
Kitchen/ 1990	–	–	–	–
Sostek/ 1987	–	–	–	–
van de Bor/ 1988	–	–	–	–
van de Bor/ 1993	–	–	–	–
van de Bor/ 2004	–	–	–	–
Whitaker/ 1996	–	–	–	–
Whitaker/ 1997	–	–	–	–
Pinto-Martin/ 1999	–	–	–	–
Hanigan/ 1991	–	–	–	–
Lowe/ 1990	–	–	–	–
Levene/ 1992	–	–	–	–
Sherlock/ 2005	–	–	–	–
Ross/ 1992	–	–	–	–
Ross/ 1996	–	–	–	–

Plus mark (+) means the criterion is established. Minus mark (–) means the criterion is not established.

Appendix Table 1 (1 of 2): Measurements and Classification of IVH

Author/ year	Times and Timeframe of Exposure Measurement
Papile/ 1983	Examined twice. 1st: performed CT between 3 and 10 days of age. 2nd: performed CT or US one week after the initial scan.
Ment/ 1985	Examined at least 3 times. Performed CT scans within the first three weeks of life. Performed US scans ≧ twice in the first eight weeks of life.
Kitchen/ 1990	Performed US several times between birth and after discharge until any ventricular dilation had either resolved or stabilized.
Sostek/ 1987	Examined four times: performed US at the time of birth, on days 1 and 7, and at discharge.
van de Bor/ 1988	Performed US at least four times. 1st: as soon as after admission. 2nd-3rd: within the first week of life. 4th: after the first week of life until discharge.
van de Bor/ 1993	Performed US at least four times. 1st: as soon as after admission. 2nd-3rd: within the first week of life. 4th: after the first week of life until discharge.
van de Bor/ 2004	Performed US at least four times. 1st: as soon as after admission. 2nd-3rd: within the first week of life. 4th: after the first week of life until discharge.
Whitaker/ 1996	Performed US 3-4 times. 1st- 3rd: 98% of the cohort was examined at least one time at 4 hours, 24 hours, and in 7 days of life. 4th: 47% of the cohort was examined between the third and fifth hospital week and/ or before discharge.
Whitaker/ 1997	Performed US 3-4 times. 1st- 3rd: 98% of the cohort was examined at least one time at 4 hours, 24 hours, and in 7 days of life. 4th: 47% of the cohort was examined between the third and fifth hospital week and/ or before discharge.
Pinto-Martin/ 1999	Performed US 3-4 times. 1st- 3rd: 98% of the cohort was examined at least once at 4 hours, 24 hours, and in 7 days of life. 4th: 47% of the cohort was examined between the third and fifth hospital week and/ or before discharge.

Hanigan/ 1991	Performed US 4 times. 1st: at 12 hours of age. 2nd-4th: at 2, 3, 7, and 14 days after birth.
Lowe/ 1990	Performed a CT brain scan at 5 to 10 days of age. Given infants with PIVH weekly CT brain scan until the hemorrhage resolved.
Levene/ 1992	Examined at least twice: weekly performed US for the first month and then every week until discharge.
Sherlock/ 2005	Performed at least one US in the first week of life, at 28 days, and prior to discharge.
Ross/ 1992	Performed US three times. 1st: at two weeks from birth. 2nd: at before discharge. 3rd: at 1 month after birth.
Ross/ 1996	Performed US three times. 1st: at two weeks from birth. 2nd: at before discharge. 3rd: at 1 month after birth.

IVH: Intraventricular Hemorrhage; CT: Computed Tomographic; US: Ultrasound; CIVH: Cerebral Intraventricular Hemorrhage; GMH: Germinal Matrix Hemorrhage; GMH/IVH: Germinal Matrix Hemorrhage and/or Intraventricular Hemorrhage; CVH: Cerebroventricular Hemorrhage; ICH: Intracerebral Hemorrhage; VD: Ventricular Dilation; PIVH: Periventricular-Intraventricular Hemorrhage; PL/VE: Parenchymal Lesion and/or Ventricular Enlargement; PVH: Periventricular hemorrhage; FT: Full Term.

Appendix Table 1 (2 of 2): Measurements and Classification of IVH

Author/ year	Classification of IVH
Papile/ 1983	CIVH; grade from I to IV. Grade I: isolated GMH. Grade II: IVH with normal ventricular size. Grade III: IVH with ventricular dilation. Grade IV: IVH with parenchymal hemorrhage.
Ment/ 1985	GMH/IVH; grade from I to IV; according to the Papile classification.
Kitchen/ 1990	CVH and ventricular dilation were defined as follows: (1) GMH: "hemorrhage, usually less than 1 cm diameter, confined to the germinal matrix." (2) IVH: "all hemorrhages within the lateral ventricles, regardless of size, but not extending into brain parenchyma." (3) ICH: "any hemorrhage extending into or originating in brain parenchymalateral or superior to the lateral ventricles." (4) VD: "width of the lateral ventricles exceeding 3 mm in the coronal view at the level of the foramen of Monro."
Sostek/ 1987	IVH; grade from I to IV; according to a modified Papile classification. Grade I: GMH without ventricular blood. Grade II: IVH in the lateral ventricles without ventricular distention. Grade III: IVH with ventricular distention. Grade IV: IVH extending into the brain parenchyma.
van de Bor/ 1988	PIVH; grade from I to IV; according to the Papile classification.
van de Bor/ 1993	PIVH; grade from I to IV; according to the Papile classification.
van de Bor/ 2004	PIVH; grade from I to IV; according to the Papile classification.
Whitaker/ 1996	Three study groups: (1) No abnormality (NA) (2) GMH/ IVH (3) PL/VE
Whitaker/ 1997	Three study groups: (1) No abnormality (NA) (2) GMH/ IVH (3) PL/VE
Pinto-Martin/ 1999	Three study groups: (1) No abnormality (NA) (2) GMH/ IVH (3) PL/VE
Hanigan/ 1991	PVH; grade from I to IV; according to the Papile classification. (1) Low-grade hemorrhages: Papile's classification Grades I & II.

	(2) High-grade hemorrhages: Papile's classification Grades III & IV.
Lowe/ 1990	PIVH; grade from I to II; according to the Papile classification.
Levene/ 1992	Five study groups: (1) Premature normal ultrasound group (2) Prolonged flare (PF) (3) GM/IVH, without parenchymal hemorrhage, but no evidence of PF. (4) Both GM/IVH and PF (5) FT Controls group
Sherlock/ 2005	IVH; grade from I to IV; according to the Papile's classification.
Ross/ 1992	Three study groups: (1) Premature normal ultrasound group (2) Subependymal and Intraventricular Hemorrhage (S/IVH) (3) FT Controls group
Ross/ 1996	Three study groups: (1) Premature normal ultrasound group (2) Subependymal and Intraventricular Hemorrhage (S/IVH) (3) FT Controls group

IVH: Intraventricular Hemorrhage; CT: Computed Tomographic; US: Ultrasound; CIVH: Cerebral Intraventricular Hemorrhage; GMH: Germinal Matrix Hemorrhage; GMH/IVH: Germinal Matrix Hemorrhage and/or Intraventricular Hemorrhage; CVH: Cerebroventricular Hemorrhage; ICH: Intracerebral Hemorrhage; VD: Ventricular Dilation; PIVH: Periventricular-Intraventricular Hemorrhage; PL/VE: Parenchymal Lesion and/or Ventricular Enlargement; PVH: Periventricular hemorrhage; FT: Full Term.

Appendix Table 2 (1 of 2): Measurements of Outcomes

Author/ Year	Outcome
Papile/ 1983	1. Overall development. 2. Neuromotor. 3. Vision and hearing.
Ment/ 1985	1. Development (IQ) 2. Development (IQ)
Kitchen/ 1990	1. Vision, hearing, and the central nervous system (i.e.: Spastic CP; Ataxic CP) 2. Psychologic assessment
Sostek/ 1987	1. Mental and motor development 2. Muscle tone, primitive reflexes, deep tendon reflexes, protective and equilibrium reactions, and range of motion.
van de Bor/ 1988	1. Neurodevelopmental outcome
van de Bor/ 1993	1. Mental development 2. Abnormalities of gross and fine motor function 3. Neurological status 4. Minor neurological dysfunction 5. Central motor deficit 6. Vision and visual functions. 7. Hearing 8. Speech and language development
van de Bor/ 2004	1. Neurodevelopmental outcome was assessed during a home visit at 5 years (including neurological status, vision and visual function, hearing, and language and speech). & Mental development. 2. School performance at 9 years of age. 3. School performance at 14 years of age.
Whitaker/ 1996	Age 6 years 1. Global cognitive outcomes 2. Specific cognitive outcomes (a) Language; (b) Short-term memory; (c) Quantitative reasoning; (d) Visual perceptual organization.
Whitaker/ 1997	Age 6 years 1. DSM-III-R psychiatric disorders; Behavior problems (i.e.: children with ADHD) 2. Intelligence 3. Motor problems
Pinto-Martin/	Age 2 years

1999	1. Neurodevelopmental status 2. "Standardized assessments of infant development"; "Examinations of vision, hearing, and CP" <u>Age 6 years</u> 1. Psychopathology 2. "Standardized assessments of general intellectual and motor functioning" <u>Age 9 years</u> 1. School performance 2. "Standardized assessments of general intellectual and motor functioning"
Hanigan/ 1991	1. <u>Neurodevelopmental outcome:</u> Ex: "Abnormalities in muscle tone, reflexes, postural reactions and motor patterns" 2. <u>Cognitive status</u> 3. <u>Developmental quotients</u>: Ex: Visual motor/ problem-solving skills, receptive language, and expressive language
Lowe/ 1990	1. Cognitive performance 2. Early readying skills 3. Visuomotor integration 4. Reading potential
Levene/ 1992	1. Motor functions 2. Vocabulary
Sherlock/ 2005	1. CP; blindness; deafness 2. Intellectual impairments (IQ score< -1 S.D.)
Ross/ 1996	1. Motor & Mental. 2. Three special cognitive abilities.
Ross/ 1992	1. "Infants' information processing" 2. "The development of memory for location of an object" 3.Motor & Mental.

Appendix Table 2 (2 of 2): Measurements of Outcomes

Author/ Year	Measurements of Outcome
Papile/ 1983	1. Bayley Scales of Infant Development (including mental (MDI) and motor (PDI) scales). 2. Neuromotor examination 3. Ophthalmologic and audiologic examinations
Ment/ 1985	1. BSID at 3, 6, 12, and 18 months. 2. SBIS and PPVT-R at 30 months.
Kitchen/ 1990	1. Diagnosed by the developmental pediatrician. 2. WPPSI (a test in English)
Sostek/ 1987	1. The Bayley Scales of Infant Development (Bayley, 1969) 2. An extensive neurologic examination based on Amiel-Tison and Grenier (1980)
van de Bor/ 1988	1. Gesell test
van de Bor/ 1993	1. A language and speech test 2. DDST 3. Part of Touwen's standardized and age-specific neurological examination 4. Observation of its presence 5. SRCMDASID 6. Physical, motor, and functional examination of vision 7. Puretone audiometry with a hand-held audiometer (Hortmann DA 323) 8. Screening test developed by Gerritsen (1989)
van de Bor/ 2004	1. Validated standardized tests (was assessed at 5 years). Registration of school attendance at 5 years of age. 2. Using questionnaires to be filled by parents (outcomes: (1) normal, (2) slow learners, and (3) special education). 3. Using questionnaires to be filled by adolescents (outcomes: (1) normal, (2) slow learners, and (3) special education).
Whitaker/ 1996	Age 6 years 1. SB fourth edition composite score (for general intellectual functioning); VABS composite score (for overall adaptive functioning). 2. (a) TOLD (for overall, receptive, expressive language); SB verbal reasoning area score (for verbal reasoning); (b) SB short-term memory area score (for short-term memory); (c) SB quantitative reasoning area score (for quantitative reasoning); (d) SB abstract visual reasoning area score (for abstract visual reasoning); VMI (for visual motor integration; TVPS (for visual perceptual skills).

Whitaker/ 1997	Age 6 years 1. DISC 2.1 P (the children's version was not used). 2. SB 3. RMPI
Pinto-Martin/ 1999	Age 2 years 1. Motor outcome measure: Bayley PDI (Bayley 1969) 2. General cognitive outcome measure: Bayley MDI (Bayley 1969) and SB3 (Terman and Merrill 1960). Age 6 years 1. Motor outcome measure: RMPI (Riley 1976) (assess fine, gross, and oral). 2. General cognitive outcome measure: SB:FE (Thorndike et al. 1986) Age 9 years 1. Motor outcome measure: RMPI (assess fine, gross, and oral). 2. General cognitive outcome measure: WISC-III (Wechsler 1991)
Hanigan/ 1991	1. The Peabody DMQs 2. A nonstandardized battery of items taken from Gesell, Cattell, and Bayley. BSID (for the 18-month visit). SBIS (for the 3-year visit).
Lowe/ 1990	1. McCSCA 2. TERA 3. BDTVMI 4. BFL
Levene/ 1992	1. TOMI (Stott et al., Henderson) 2. WPPSI 3. A number of measures of basic literacy, language, representational and mathematical concepts and skills (Tizard et al.). 4. A checklist of behavioral statements.
Sherlock/ 2005	1. Diagnosed by pediatricians; Movement ABC. 2. WISC-III; Several special cognitive tests (including VCI, POI, FDI, PSI, TOL, RCF, WRAT3)
Ross/ 1996	1. The Bayley Scales of Infant Development (Bayley, 1969) 2. A habituation/novelty preference task (Janowsky, 1985); an object displacement task; an object discrimination reversal task

Ross/ 1992	1. A habituation/novelty preference task (Janowsky, 1985) 2. The AB object permanence task (Diamond, 1985) 3. The Bayley Scales of Infant Development (Bayley, 1969): Bayley MDI, PDI.

N: Number; DDST: Denver Developmental Screening Test; SRCMDASID: the Standard Recording of Central Motor Deficit and Associated Sensory and Intellectual Deficit; Movement ABC: the Movement Assessment Battery for Children; WISC-III: the Wechsler Intelligence Scale for Children-Third Edition; VCI: Verbal Comprehension Index; POI: Perceptual Organization Index; FDI: Freedom from Distractibility Index; PSI: Processing Speed Index; TOL: Tower of London; RCF: Rey Complex Figure; WRAT3: Wide Range Achievements Test; CP: Cerebral Palsy; Bayley PDI: Bayley Psychomotor Development Index; Bayley MDI: Bayley Mental Developmental Index; SB3: the Stanford Binet Third Edition, Form L-M; RMPI: the Riley Motor Problems Inventory; SB:FE: the Composite Index of the Stanford Binet: Fourth Edition; WISC-III: the Full-Scale IQ from the Wechsler Intelligence Scales for Children- Third Edition; DISC 2.1 P: the Diagnostic Interview Schedule for Children-Parent version 2.1P; SB: the Composite Index of the Stanford Binet; VABS: Vineland Adaptive Behavior Scale; TOLD: Test of Language Development; VMI: the Developmental Test of Visual Motor Integration; TVPS: the Test of Visual Perceptual Skills; The Peabody DMQs: the Peabody Developmental Motor Quotients; BSID: the Bayley Scales of Infant Development; SBIS: the Stanford-Binet Intelligence Scale; WPPSI: the Wechsler Preschool and Primary Scales of Intelligence; K-ABC: the Gestalt Closure subtest from the Kaufman Assessment Battery for Children; HVOT: the Hooper Visual Organization Test; GP test: the Grooved Pegboard test; ROCF: the Rey-Osterreith Complex Figure; McCSCA: the McCarthy Scales of Children's Ability; TERA: the Test of Early Reading Ability; BDTVMI: the Beery Developmental Test of Visual-Motor Integration; BFL: Benton Finger Localization; SBISFE: the Stanford-Binet Intelligence Scale, Fourth Edition; VR: Verbal Reasoning; A/VR: Abstract/ Visual Reasoning; STMSSAS: Short Term Memory Subscale Standard Age Scores; SICD: the Sequenced Inventory of Communication Development; ECA: Expressive Communication Age; RCA: Receptive Communication Age; TOMI: the Test of Motor Impairment; WPPSI: the vocabulary subscale of the Wechsler Preschool and Primary Scale of Intelligence; SBIS: the Stanford-Binet Intelligence Scale; PPVT-R: the Peabody Picture Vocabulary Test-Revised.

BIBLIOGRAPHY

1. van de Bor M and Ouden L. School performance in adolescents with and without periventricular-intraventricular hemorrhage in the neonatal period. Sem. Perin. Vol 28, No 4, 2004

2. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. J Pediatr. 1978 Apr;92(4):529-34.

3. Dykes FD, Lazzara A, Ahmann P, Blumenstein B, Schwartz J, Brann AW. Intraventricular hemorrhage: a prospective evaluation of etiopathogenesis. Pediatrics. 1980 Jul;66(1):42-9.

4. Ahmann PA, Lazzara A, Dykes FD, Brann AW Jr, Schwartz JF. Intraventricular hemorrhage in the high-risk preterm infant: incidence and outcome. Ann Neurol. 1980 Feb;7(2):118-24.

5. Fitzhardinge PM, Flodmark O, Fitz CR, Ashby S. The prognostic value of computed tomography of the brain in asphyxiated premature infants. J Pediatr. 1982 Mar;100(3):476-81.

6. Shankaran S, Slovis TL, Bedard MP, Poland RL. Sonographic classification of intracranial hemorrhage. A prognostic indicator of mortality, morbidity, and short-term neurologic outcome. J Pediatr. 1982 Mar;100(3):469-75.

7. Ment LR, Scott DT, Ehrenkranz RA, Rothman SG, Duncan CC, Warshaw JB. Neonates of less than or equal to 1,250 grams birth weight: prospective neurodevelopmental evaluation during the first year post-term. Pediatrics. 1982 Aug;70(2):292-6.

8. Rennie JM. The immature brain. In: Rennie JM, editor. Neonatal cerebral ultrasound. Cambridge University Press; 1997. 124.

9. Escobar GJ, Littenberg B, Petitti DB. Outcome among surviving very low birthweight infants: a meta-analysis. Arch Dis Child. 1991 Feb;66(2):204-11.

10. Pinto-Martin JA, Whitaker AH, Feldman JF, Van Rossem R, Paneth N. Relation of cranial ultrasound abnormalities in low-birthweight infants to motor or cognitive performance at ages 2, 6, and 9 years. Dev Med Child Neurol. 1999 Dec;41(12):826-33.

11. Aylward GP, Pfeiffer SI, Wright A, Verhulst SJ. Outcome studies of low birth weight infants published in the last decade: a meta analysis. J Pediatr. 1989 Oct;115(4):515-20.

12. Whitaker A. H., Feldman J. F., J. A. Pinto-Martin, N. Paneth. Neonatal Cranial Ultrasound Abnormalities in LBW infants: relation to cognitive outcomes at six years of age. Pediatrics, vol 98, No. 4. 1996

13. Paneth N., Rudelli R., Kazam E., Monte W. Brain damage in the preterm infant. 1994 Mac Keith Press

14. Whitaker A.H., Rossem R. V., J. A. Pinto-Martin, N. Paneth. Psychiatric Outcomes in low-birth-weight children at age 6 years: relation to neonatal cranial ultrasound abnormalities Arch G Psy vol 54, sep 1997

15. Hack M, Flannery DJ, Schluchter M, Cartar L, Borawski E, Klein N. Outcomes in young adulthood for very-low-birth-weight infants. N Engl J Med. 2002 Jan 17;346(3):149-57.

16. Hack M, Taylor HG, Klein N, Mercuri-Minich N. Functional limitations and special health care needs of 10- to 14-year-old children weighing less than 750 grams at birth. Pediatrics. 2000 Sep;106(3):554-60.

17. Hack M, Fanaroff AA. Outcomes of children of extremely low birthweight and gestational age in the 1990's. Early Hum Dev. 1999 Jan;53(3):193-218.

18. Saigal S, Szatmari P, Rosenbaum P, Campbell D, King S. Cognitive abilities and school performance of extremely low birth weight children and matched term control children at age 8 years: a regional study. J Pediatr. 1991 May;118(5):751-60.

19. Whitfield MF, Grunau RV, Holsti L. Extremely premature (< or = 800 g) schoolchildren: multiple areas of hidden disability. Arch Dis Child Fetal Neonatal Ed. 1997 Sep;77(2):F85-90.

20. Volpe JJ. Neurology of the newborn Fourth Edition. 2001

21. Rakic PT. Embryonic development of the pulvinar-LP complex in man. In: Cooper IS, Riklan M, Rakic PT (Eds.) The pulvinar-LP complex. 1974, Springfield, IL: Thomas CC: 3-35

22. Rakic PT, Sidman RL. Telencephalic origin of pulvinar neurons in the fetal human brain. Zeitschrift fur Anatomie and Entwicklungsgeschichte. 1969; 129: 53-82

23. Hambleton G, Wigglesworth JS. Origin of intraventricular haemorrhage in the preterm infant. Arch Dis Child. 1976 Sep;51(9):651-9.

24. Larroche JC. Intraventricular hemorrhage in the premature neonate. In Korobkin R, Guilleminault C, editors: Advances in perinatal neurology, vol 1, New York, 1979, SP Medical and Scientific Books.

25. Yakovlev PI, Rosales RK. Distribution of the terminal hemorrhages in the brain wall in stillborn premature and nonviable neonates. In Angle CR, Bering EA Jr, editors: Physical trauna as an etiologic agent in mental retardation, Washington DC, 1970, US Government Printing Office.

26. Gruenwald P. Subependymal cerebral hemorrhage in premature infants, and its relation to various injurious influences at birth. Am J Obstet Gynecol. 1951 Jun;61(6):1285-92.

27. Leech RW, Kohnen P. Subependymal and intraventricular hemorrhages in the newborn. Am J Pathol. 1974 Dec;77(3):465-75.

28. Rorke LB. Pathology of perinatal brain injury, New York, 1982, Raven Press.

29. Paneth N. Classifying brain damage in preterm infants. J Pediatr. 1999 May;134(5):527-9.

30. Jeffrey JN, Terrie EI. Imaging perinatal brain injury in premature infants. Semin Perinatol. 2004;28:433-443

31. McArdle CB, Richardson CJ, Hayden CK, Nicholas DA, Crofford MJ, Amparo EG. Abnormalities of the neonatal brain: MR imaging. Part I. Intracranial hemorrhage. Radiology. 1987 May;163(2):387-94.

32. Haddad J, Constantinesco A, Brunot B. Single photon emission computed tomography of the brain perfusion in neonates. In Haddad J, Christmann D, Messer J, editors: Imaging techniques of the CNS of the neonates, New York, 1991, Springer-Verlag.

33. Zuerrer M, Martin E, Boltshauser E. MR imaging of intracranial hemorrhage in neonates and infants at 2.35 Tesla. Neuroradiology. 1991;33(3):223-9.

34. Barkovich AJ. Pediatric neuroimaging, ed 2, New York, 1995, Raven Press.

35. Krishnamoorthy KS, Fernandez RA, Momose KJ, DeLong GR, Moylan FM, Todres ID, Shannon DC. Evaluation of neonatal intracranial hemorrhage by computerized tomography. Pediatrics. 1977 Feb;59(2):165-72.

36. Scott WR, New PF, Davis KR, Schnur JA. Computerized axial tomography of intracerebral and intraventricular hemorrhage. Radiology. 1974 Jul;112(1):73-80.

37. Pevsner PH, Garcia-Bunuel R, Leeds N, Finkelstein M. Subependymal and intraventricular hemorrhage in neonates. Early diagnosis by computed tomography. Radiology. 1976 Apr;119(1):111-4.

38. Volpe JJ. Neonatal intracranial hemorrhage. Pathophysiology, neuropathology, and clinical features. Clin Perinatol. 1977 Mar;4(1):77-102.

39. Rumack CM, McDonald MM, O'Meara OP, Sanders BB, Rudikoff JC. CT detection and course of intracranial hemorrhage in premature infants. AJR Am J Roentgenol. 1978 Sep;131(3):493-7.

40. Lee BC, Grassi AE, Schechner S, Auld PA. Neonatal intraventricular hemorrhage: a serial computed tomography study. J Comput Assist Tomogr. 1979 Aug;3(4):483-90.

41. Albright L, Fellows R. Sequential CT scanning after neonatal intracerebral hemorrhage. AJR Am J Roentgenol. 1981 May;136(5):949-53.

42. Siegel MJ, Patel J, Gado MH, Shackelford GD. Cranial computed tomography and real-time sonography in full-term neonates and infants. Radiology. 1983 Oct;149(1):111-6.

43. O'Shea TM, Counsell SJ, Bartels DB, Dammann O. Magnetic resonance and ultrasound brain imaging in preterm infants. Early Hum Dev. 2005 Mar;81(3):263-71.

44. Tonukari NJ. Searching the online biomedical literature from developing countries. African J of Biotechnology. 2005 Aug; 4 (8): 758-762

45. Detmer WM. Medline on the web: ten questions to ask when evaluating a web based

service. Internet Working Group Newsletter 1997; 3: 11-13.

46. Vincent B, Vincent M, Ferreira CG. Making PubMed searching simple: learning to retrieve medical literature through interactive problem solving. Oncologist. 2006 Mar;11(3):243-51.

47. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pontius JU, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D39-45.

48. Day J. The quest for information: a guide to searching the Internet. J Contemp Dent Pract. 2001 Nov 15;2(4):33-43.

49. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ. 1994 Nov 12;309(6964):1286-91

50. Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, Adams A, Mishra SI. Searching biomedical databases on complementary medicine: the use of controlled vocabulary among authors, indexers and investigators. BMC Complement Altern Med. 2003 Jul 7;3:3. Epub 2003 Jul 7.

51. Coletti MH, Bleich HL. Medical subject headings used to search the biomedical literature. J Am Med Inform Assoc. 2001 Jul-Aug;8(4):317-23.

52. Dunning J, Prendergast B, Mackway-Jones K. Towards evidence-based medicine in cardiothoracic surgery: best BETS. Interactive Cardiovascular and Thoracic Surgery 2003; 2: 405-409

53. Pappas DE, Owen Hendley J. Otitis media. A scholarly review of the evidence. Minerva Pediatr. 2003 Oct; 55(5): 407-14.

54. Harrison J. Designing a search strategy to identify and retrieve articles on evidence-based health care using MEDLINE. Health Libr Rev. 1997 Mar;14(1):33-42.

55. Stewart MG, Kuppersmith RB, Moore AS. Searching the medical literature on the Internet. Otolaryngol Clin North Am. 2002 Dec; 35(6): 1163-74, v-vi.

56. Weed DL. Methodologic guidelines for review papers. J Nation. Canc. Instit. 1997 Jan; 89(1): 6-7

57. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. Int J Epidemiol. 1999 Feb;28(1):1-9.

58. Evans D. Database searches for qualitative research. J Med Libr Assoc. 2002 Jul;90(3):290-3.

59. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA. 1994 Apr 13;271(14):1103-8.

60. Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA. 1990 Mar 9;263(10):1385-9.

61. Dickersin K. How important is publication bias? A synthesis of available data. AIDS Educ Prev. 1997 Feb;9(1 Suppl):15-21.

62. Rosenthal R. The "file-drawer problem" and tolerance for null results. Psychol Bull. 1979;86:638-41

63. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? Lancet. 1993 Feb 13;341(8842):418-22.

64. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet. 1996 Jun;347(9017):1713-27

65. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. Stat Med. 1995 Oct 15;14(19):2057-79.

66. Breslow RA, Ross SA, Weed DL. Quality of reviews in epidemiology. Am J Public Health. 1998 Mar;88(3):475-7.

67. Weed DL, Gorelic LS. The practice of causal inference in cancer epidemiology.

Cancer Epidemiol Biomarkers Prev. 1996 Apr;5(4):303-11.

68. Weed DL. On the use of causal criteria. Int J Epidemiol. 1997 Dec;26(6):1137-41.

69. Hill AB. The environment and disease: association or causality? Proc R Soc Med. 1965;58:295-300.

70. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology--a basic science for clinical medicine. Boston: Little, Brown, 1985.

71. Holland PW, Statistics and causal inference. J Am Statist Assoc. 1986;81:945-60.

72. Peterson HB, Kleinbaum DG. Interpreting the literature in obstetrics and gynecology. 1. Key concepts in epidemiology and biostatistics. Obstet Gynecol 1991;78:710-7.

73. Nicol-Smith L. Causality, menopause, and depression: a critical review of the literature. BMJ. 1996 Nov 16;313(7067):1229-32.

74. Elwood JM. Causal relationships in medicine. A practical system for critical appraisal. Oxford medical publications. 1988

75. Papile LA, Munsick-Bruno G, Schaefer A. Relationship of cerebral intraventricular hemorrhage and early childhood neurologic handicaps. J Pediatr. 1983 Aug;103(2):273-7.

76. Ment LR, Scott DT, Ehrenkranz RA, Duncan CC. Neurodevelopmental assessment of very low birth weight neonates: effect of germinal matrix and intraventricular hemorrhage. Pediatr Neurol. 1985 May-Jun;1(3):164-8.

77. Sostek AM, Smith YF, Katz KS, Grant EG. Developmental outcome of preterm infants with intraventricular hemorrhage at one and two years of age. Child Dev. 1987 Jun;58(3):779-86.

78. Hanigan WC, Morgan AM, Anderson RJ, Bradle P, Cohen HS, Cusack TJ, Thomas-McCauley T, Miller TC. Incidence and neurodevelopmental outcome of periventricular hemorrhage and hydrocephalus in a regional population of very low birth weight infants. Neurosurgery. 1991 Nov; 29(5):701-6.

79. Ross G, Boatright S, Auld PA, Nass R. Specific cognitive abilities in 2-year-old children with subependymal and mild intraventricular hemorrhage. Brain Cogn. 1996

Oct;32(1):1-13.

80. Boss G, Tesman J., Auld PAM, Nass R. Effects of Subependymal and Mild Intraventricular Lesions on Visual Attention and Memory in Premature Infants. Developmental Psychology 1992. Vol 28. No. 6. 1067-1074

81. Lowe J., Papile L.A. Neuridevelopmental Performance of Very-Low-Birth-Weight Infants With Mild Periventricular, Intraventricular Hemorrhage. Outcome at 5 to 6 Years of Age. AJDC Vol 144, Nov. 1990

82. van de Bor M, Verloove-Vanhorick SP, Baerts W, Brand R, Ruys JH. Outcome of periventricular-intraventricular hemorrhage at 2 years of age in 484 very preterm infants admitted to 6 neonatal intensive care units in The Netherlands. Neuropediatrics. 1988 Nov;19(4):183-5.

83. van de Bor M, Ens-Dokkum M, Schreuder AM, Veen S, Brand R, Verloove-Vanhorick SP. Outcome of periventricular-intraventricular haemorrhage at five years of age. Dev Med Child Neurol. 1993 Jan;35(1):33-41.

84. Sherlock R. L., Anderson P. J., Doyle L. W. Neurodevelopmental sequelae of intraventricular haemorrhage at 8 years of age in a regional cohort of ELBW/ very preterm infants. E.H.D. (2005)

85. Kitchen WH, Ford GW, Rickards AL, Doyle LW, Kelly E, Murton LJ. Five-year outcome of infants of birthweight 500 to 1500 grams: relationship with neonatal ultrasound data. Am J Perinatol. 1990 Jan;7(1):60-5.

86. Levene M, Dowling S, Graham M, Fogelman K, Galton M, Phillips M. Impaired motor function (clumsiness) in 5 year old children: correlation with neonatal ultrasound scans. Arch Dis Child. 1992 Jun;67(6):687-90

87. Hack M, Taylor HG, Klein N, Eiben R, Schatschneider C, Mercuri-Minich N. School-age outcomes in children with birth weights under 750 g. N Engl J Med. 1994 Sep 22;331(12):753-9.

88. Pagano M., Gauvreau K. Principles of biostatistics. Duxbury press, 1993.

89. Bhopal R. Concepts of epidemiology- an integrated introduction to the ideas, theories, principles and methods of epidemiology. Oxford University Press. 2002.

90. Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. J Epidemiol Community Health. 2004 Jul;58(7):538-45.

91. Dickersin K. Systematic reviews in epidemiology: why are we so far behind? Int J Epidemiol. 2002 Feb;31(1):6-12.

92. Mulrow CD. Rationale for systematic reviews. BMJ. 1994 Sep 3;309(6954):597-9.

93. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988;138:697-703.

94. Cohen, J. Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum. 1988

95. Scargle JD. Publication bias: the "file-drawer" problem in scientific inference. J. of Scientific Exploration. 2000; 14(1): 91-106.

96. http://sportsci.org/resource/stats/index.html